FILED

IN THE CIRCUIT COURT OF THE NINETEENTH JUDICIAL CIRCUIT JAN 2 6 2024
LAKE COUNTY, ILLINOIS

Erin Cartwright Weinstein
CIRCUIT CLERK

PEOPLE OF THE STATE OF ILLINOIS *Respondent-Plaintiff*	)	No. 00 CF 1920
v.	)	
MARVIN T. WILLIFORD *Petitioner-Defendant*	)	Hon. Daniel Shanes, Judge Presiding

NOTICE OF MOTION

TO: Lake County State's Attorney's Office
Eric F. Rinehart
18 N. County Street
Waukegan, IL 60085
erinehart@lakecountyil.gov

Please take notice that on February 1, 2024, I will appear in front Honorable Judge Shanes to present Williford's **Motion for Leave to File An Amended Successive Petitioner for Post-Conviction Relief**, to be subsequently filed in the above-entitled cause with the Clerk of the Circuit Court of Lake County, 18 N. County Street, Waukegan, IL 60085.

Dated: January 25, 2024

/s/ David B. Owens
Atty for Marvin Williford

David B. Owens
The Exoneration Project
311 N. Aberdeen St., 3rd Floor
Chicago, IL 60607
(312)789-4955
david@exonerationproject.org
Atty No. 44407

IN THE CIRCUIT COURT OF LAKE COUNTY, ILLINOIS

FILED
JAN 2 6 2024
CIRCUIT CLERK

PEOPLE OF THE STATE OF ILLINOIS, Respondent	)	
v.	)	No. 00 CF 1920
MARVIN WILLIFORD, Petitioner	)	Hon. Chief Judge Shanes, Presiding

MOTION FOR LEAVE TO FILE AMENDED SUCCESSIVE PETITION <u>FOR POST-CONVICTION RELIEF</u>

Now comes Petitioner, MARVIN WILLIFORD, by and through his attorneys, The Exoneration Project, and hereby respectfully submits this Motion for Leave to File An Amended Successive Petition for Post-Conviction Relief. In support, Petitioner states:

INTRODUCTION

Centered on an unreliable stranger's misidentification, Marvin Williford stands wrongfully convicted of murder relative to the January 2000 home invasion of Delwin Foxworth.

Williford is not alone in being wrongfully convicted based upon a misidentification. "Studies have shown that erroneous identification accounted for as much as 85% of the convictions later exonerated by DNA testing." <u>People v. Lerma</u>, 2014 IL App (1st) 121880, ¶ 39 (citing Jacqueline McMurtie, *The Role of the Social Sciences in Preventing Wrongful Convictions*, 42 AM. CRIM. L. REV. 1271, 1271-75 (2005)). Indeed, as the Illinois Supreme Court has recognized, that "'eyewitness misidentification is now the single greatest source of wrongful convictions in the

United States, and responsible for more wrongful convictions than all other causes combined.'"
People v. Lerma, 2016 IL 118496, ¶ 24 (quoting State v. Dubose, 285 Wis.2d 143 (2005)).

In 2015, Williford filed a post-conviction petition that included substantial evidence of his innocence. After a hearing, that petition was denied. The denial was affirmed on appeal but remains subject to federal *habeas corpus* litigation.

In four general ways, new developments since 2015 further confirm that Williford is innocent, that this is a case of mistaken identification, and—at minimum—demonstrate that Williford should be permitted to seek post-conviction relief.

First, as it relates to Williford's claim of actual innocence, after this Court and the Court of Appeals issued decisions denying the petition, the Illinois Supreme Court in People v. Robinson, 2020 IL 123849, confirmed that dicta suggesting "total vindication" or "exoneration" is required to prevail on an actual innocence claim is mistaken. The dismissal of Williford's prior petition involved citation to these sorts of authorities. In other words, the standard applied assessing aspects of Williford's prior petition was too high. On the appropriate standard, Williford should have prevailed on the basis of the substantial evidence in his prior petition because he was not required to prove, beyond all metaphysical doubt, that he was innocent. Rather, he only needed to demonstrate that the new evidence would probably change the result on retrial—a standard that does not require evidence be even completely dispositive of an issue to warrant relief. Id. ¶56 (citing People v. Coleman, 2013 IL 113307, ¶ 97).

In other words, even if there were no new evidence discovered since 2014, Williford is entitled under Robinson to have his actual innocence claims cumulatively evaluated on the governing legal standard and without any reference to language derived from *dicta* in now-overruled cases.

That said, there is new evidence in the case further showing that Williford should respectfully be permitted to pursue his claims anew and showing that he is entitled to post-conviction relief.

Second, this case is one where the trial "identification" followed: (1) a photographic show-up nearly 6 months after the crime; (2) a procedure of reminding the witness about the prior identification and showing her the same picture again, albeit with a few others, some two-and-a-half years after the crime; and (3) a third procedure, a 6-pack lineup over three years after the crime, where the witness hesitated substantially before settling on the picture of Williford. (Williford is the only person whose picture was presented in all three instances).

New, important developments in the social science concerning memory and cognition confirm that the "identification" evidence of Williford was unreliable and should not have been admitted. Studies published after 2015—including a 2020 consensus "white paper" document summarizing the science in the field—address (1) how to evaluate the contamination of a witness's identification made *after* an unfair procedure, and (2) evidence of unreliability by the lack of confidence Conners expressed at the time of the third procedure. These studies confirm that it is *impossible*, as a scientific matter, for Conners to have reliably identified the perpetrator after having been subjected to the multiple unfair pretrial procedures, particularly given the doubt she expressed at the third procedure. Despite the State's prior reasoning to the contrary—*i.e.*, that Conners could have somehow made a constitutionally-adequate, untainted identification despite the contaminating impacts of the (unfair) pretrial confrontations—the new social science shows that is not the case. Relief is warranted on this basis and, at minimum, Williford should be entitled to present this evidence in a successive petition.

Third, there has been additional DNA testing and analysis since 2015. That evidence puts the case in a new light and shows, further, that Williford was misidentified. This includes: (1) further testing by the NIRCL lab in 2020; and (2) analysis by DNA expert Karl Reich based upon the 2020 testing and in view of the entire case. These results show the complete exclusion of Williford based upon highly sensitive DNA techniques now available in ways more probative and sophisticated than prior testing. In addition: (3) there has been genealogical examination of the DNA by a private lab in reports produced to the defense from 2018, 2019, 2020, and 2022. These results provide powerful, previously unknown information about the identity of the person whose DNA was on the murder weapon. In eerily accurate and unexpected fashion, these new results are consistent with how Conners described the assailant the night of the crime as a "yellow skinned." This was an issue at trial, and the defense argued in the original proceedings Williford was not the "yellow-skinned" man. The new DNA genealogical analysis adds tremendous forensic support for this argument and was not corroborated in any fashion by any of the (very limited) forensic evidence in the original trial. To be clear: this DNA revelation is significant, even without reference to the fact that the DNA hit is to the same person who killed a child in Lake Couty years before the Foxworth home invasion.

Fourth, and still related to the misidentification of Williford, there is new material evidence that was suppressed by the North Chicago Police Department officers who investigated Williford, in violation of due process under *Brady v. Maryland*, 373 U.S. 83 (1963), *Giglio v. United States*, 405 U.S. 150 (1972), and other authorities. Bearing on why Williford was the only suspect ever pursued in the case and why the officers conducted a highly-suggestive photographic show-up of Williford, the evidence shows that the officers lied in their reports and then lied at trial about what took place—facts they did not disclose in violation of due process.

The new evidence is summarized more fully below. In short, this evidence shows that Williford should prevail on his post-conviction claims. And, at minimum, that Williford should be given leave to file a successive post-conviction petition so all of this can be presented in a subsequent filing. Williford therefore respectfully requests leave to file a successive post-conviction petition presenting legal claims based upon the matters described below.

BACKGROUND

Williford has steadfastly maintained his innocence of this crime, including at sentencing R.1298-99,[1] and in response to rejected plea deals that cost him opportunities to be free.

A. Relevant Procedural History

The main issue at trial was the identity of the lead perpetrator (called "T"), who beat Foxworth with a 2x4, doused him with gasoline, and fled when the room caught fire. *See* R.638-39 (identification); *id.* at 691, 698-99, 707-09 (addressing identification issues). As part of the effort to show the identification was in fact a misidentification, at trial, the defense pointed out that Conners had said the perpetrator, T, was a yellow-skinned man the night of the crime and emphasized that there was no DNA evidence pointing to Williford as such a person. *E.g.*, R.609, 616, 683-84, 686 & 696. Williford testified he was innocent, and not the person Conners had identified as committing the crime. R.1030-33. As with trial, the dispute about identity occupied a significant part of closing statements. *See, e.g.*, R.1187-89, 1191-94.

Available evidence also illustrates that the jury, during its deliberations, must have been considering the issue of identity. The jury asked one question, inquiring whether the State's exhibit, No. 75 that was sent to the jury room, was "the single picture that Detective Lawrence Wade showed Conners in June of 2000." R. 1242. Williford was convicted. *Id.* at 1243-45.

[1] Citations to the Record of Proceedings are designated "R."

It is **undisputed** that Conners' identification testimony at trial—which must have been the subject of the jury's lone question in some way—followed the following events:

(1) **June 2000**: Conners viewed a single picture of Williford on her way into the grand jury some five months after the incident, R.990-91 (Detective Wade Testimony).

(2) **September 2002**: Conners was reminded of the show-up and re-shown the same picture of Williford, Exhibit 1 (Holderbaum 2002 Report);

(3) **February 2003:** Connors was subjected to a third identification procedure including Williford where she said two photographs looked like T, but eventually selected Williford's picture, Exhibit 2 (Holderbaum 2003 Report); R. 699; *see* R.708-10.

A timeline looks like this:

Date	Event	Citation
Jan. 2000	Home invasion of Foxworth	R.468
June 2000	Conners shown single picture heading into grand jury	R.990-91
Sept. 2002	Conners reminded of prior procedure and shown same picture again in a group of photos she "thumbed through"	Ex. 1.[2]
Feb. 2003	Conners shown photo array with new picture of Williford and equivocates between two pictures she says both looked like perpetrator but ultimately selects Williford	Ex. 2.
Aug. 2004	Williford "identified" by Conners at trial	R.638-39

[2] There is no dispute that Conners was subject to a second identification procedure in September 2002. The relevant police report was produced to Williford from the State.

Williford's conviction was affirmed on appeal. Williford filed a timely post-conviction petition, which was incorrectly dismissed. The appellate court reversed. People v. Williford, No. 2-08-0068 (2 Dist. 2009) (Williford II).

On remand, in a November 3, 2015 filing, Williford's petition was amended to a Petition for Relief From Judgment and Amended Post-Conviction Petition. C.1090.[3]

That matter went to an evidentiary hearing, after which this Court denied relief. *See* Order Denying Relief (May 18, 2018), C.1860-1900. The denial was affirmed on appeal. *See* People v. Williford, 2020 IL App (2d) 180479-U ("Williford III").

Williford timely filed a federal *habeas corpus* petition in the Northern District of Illinois, Williford v. Brannon-Dortch, 1:21-cv-01661 (N.D. Ill), which remains pending but has been stayed subject to the resolution of these state-court proceedings.

B. New Material Decided, Developed, Created, or Unveiled Since 2015

Four categories of new material—warranting the filing of a successive post-conviction petition—are summarized below:

First, in its prior decisions rejecting Williford's actual innocence claim, this Court and the Court of Appeals both relied to some extent on *dicta* from People v. Barnslater, 373 Ill. App. 3d 512 (2007), and People v. Collier, 387 Ill. App. 3d 630, 636 (2008); *see* Exhibit 3, at 39 (citing Barnslatter and Collier); Williford III, 2020 IL App (2d) 180479-U, ¶34 (same). Those cases reason that evidence of actual innocence must completely exonerate a defendant and essentially eliminate any metaphysical possibility they could have possibly committed the crime in order to state an actual innocence claim for post-conviction relief. Id.

[3] Given its size, the citation here is to the Common Law Record used on appeal as "C."

In People v. Robinson, 2020 IL 123849, the Illinois Supreme Court overruled these aspects of Barnslater and Collier. In so doing, and in requiring all aspects of an actual innocence claim be addressed under People v. Coleman, 2013 IL 113307 and People v. Washington, 171 Ill. 2d 475, 489 (1996), Robinson held it was error to "employ[] a standard that requires total vindication or exoneration to support a claim of actual innocence." 2020 IL 123849, ¶ 55. Instead, "the new evidence supporting an actual innocence claim need not be entirely dispositive to be likely to alter the result on retrial" because "conclusive-character element requires only that the petitioner present evidence that places the trial evidence in a different light and undermines the court's confidence in the judgment of guilt." Id. ¶56 (citing Coleman, 2013 IL 113307, ¶97).

Robinson requires a fresh analysis of the new evidence submitted in the 2015 amended post-conviction petition, and additional evidence introduced during the prior third-stage hearing, even if there were no new evidence to add to the mix.[4]

But, there is new evidence.

Second, new developments in the field of eyewitness identifications have occurred since the 2015 petition was filed. This presents new evidence in support of Williford's actual innocence and due process claims. A trio of papers—sometimes referred to as a new scientific "consensus"—concerning multiple identification procedures provide significant developments in the field. Perhaps most significant, in 2020, the American Psychological Association and American Psychological-Law Society issued a "White Paper" concerning the scientific consensus among social scientists about issues related to memory and identification and addressing the

[4] That evidence, which is voluminous, is incorporated by reference here. *See generally* C.1090-1674 (Post-Conviction Petition and Exhibits); Exhibits Submitted in the Post-Conviction Hearing; Post-Conviction Hearing Record on Appeal. If it would assist the Court, Williford would gladly provide the Court with a USB memory stick with these materials.

"serious concerns about the potential unreliability of eyewitness identification evidence." Gary L. Wells, et al., *Policy and Procedure Recommendations for the Collection and Preservation of Eyewitness Identification Evidence*, 44 L. & HUM. BEHAVIOR 3, 4 (2020), attached as Exhibit 3. Among other things, and relevant here, the 2020 White Paper discusses why repeated identification procedures should be avoided and how an initial, unfair procedure can corrupt a subsequent identification. *Id.* at 25-26. Indeed, the White Paper details why it is the fairness (or unfairness) of the first identification procedure that is most relevant for evaluating the reliability of an identification, and that the "importance of focusing on the first identification cannot be emphasized strongly enough," as "[a]ny subsequent identification test with that same eyewitness and that same suspect is contaminated by the eyewitness's experience on the initial test." *Id.*

These issues were the specific focus of another study published by multiple social scientists in 2021 concerning how and why high-confidence identifications made in a courtroom are irretrievably corrupted by prior, unfair identification procedures (which undisputedly happened here). *See* Wixted et al., *Test A Witness's Memory of a Suspect Only Once*, 22 PSYCH. SCI. IN THE PUB. INT. 1S (2021), attached as Exhibit 4. In the same vein, a 2017 study delves into the relationship between confidence and identification accuracy—including a new synthesis on why, in the circumstances here, Conners' purported confidence was unreliable. John T. Wixted & Gary L. Wells, *The Relationship Between Eyewitness Confidence and Identification Accuracy: A New Synthesis*, 18 PSYCH. SCI. IN THE PUB. INT. 10 (2017), attached as Exhibit 5. These scientific studies provide fresh analysis showing that, in this case, the Court's prior conclusion that the science of eyewitness identification and memory does not impact analysis of Conners' misidentification was mistaken.

That Williford should at least be given the opportunity to present this evidence in a successive petition on the facts of this case—where identity was at issue at trial, there is no dispute that an unfair first identification procedure occurred, and where repeated procedures occurred before trial—is evident in the fact that this exact new scientific evidence was recently used to secure post-conviction relief and a declaration of innocence for Miguel Solorio, a defendant in California. *See* Deborah Lohse, *NCIP Wins Exoneration of Miguel Solorio, After 25 Years Wrongfully Behind Bars*, Santa Clara School of Law (Nov. 16, 2023), https://law.scu.edu/news/ncip-wins-exoneration-of-miguel-solorio-after-25-years-wrongfully-behind-bars/. In that case, the District Attorney in Los Angeles wrote that a "[n]ew documentable scientific consensus emerged in 2020 that a witness's memory for a suspect should be tested only once, as even the test itself contaminates the witness's memory." *Id.*[5]

Third, there is additional new DNA evidence. Before discussing this evidence, it is important to emphasize that DNA evidence does not stand alone but relates directly to casting doubt, from another angle, on the "identification" testimony presented in this case. The DNA evidence puts the flawed identification in an entirely different light than the original proceedings.

As mentioned, the new DNA evidence itself involves three parts: (1) DNA testing from the Northern Illinois Regional Crime Lab (NIRCL); (2) DNA Analysis conducted by Dr. Karl Reich following the NIRCL analysis, and (3) familial genetic testing conducted by a private company.

---

[55] While neither this Court or the Lake County' State's Attorney's Office are of course not bound to follow this statement, the fact that it was made is itself new evidence at least demonstrating that Wil-liford's claims are worth of being heard. Counsel for Willford in fact provided these three articles to the Lake County State's Attorney's Office, to former ASA Keviin Malia in early 2023 and then again to SA Rinehart in August 2023.

1. Based upon sensitive DNA analysis—far more sensitive at examining mixtures than the original trial evidence—a new report from the Northern Illinois Regional Crime lab **excludes** Williford from the DNA recovered from the 2x4 that was used to beat Foxworth by the perpetrators during the home invasion. See Exhibit 6.

2. To contextualize the overall DNA picture, Dr. Reich confirms that the exclusion of Williford occasioned by the 2020 testing—and throughout—is substantially more probative, from a scientific perspective, than the evidence presented at trial and even the prior 2014 testing. Exhibit 7 (Reich Report 2024). As previously presented, the 2x4 includes mixtures with Foxworth's blood that involve 2 people—"Stain D" and "Unidentified Male #1" from the Holly Staker murder. See Exhibit 8 (DNA Reports 13-19); Exhibit 9 (Reich Affidavit, 2015). In totality, and in light of the new 2020 DNA testing, as Dr. Reich has explained, the continued and additionally probative exclusion of Williford, when juxtaposed with the other affirmative DNA on the 2x4, which includes at least two other people, corroborates Williford's claim of innocence. Ex. 7.

3. Finally, genealogical DNA analysis conducted at a private lab following the evidentiary hearing in this case sheds substantial light on the fact that Williford was not the perpetrator, and instead provides substantial support for the contention—which a reasonable jury could easily credit—that the perpetrator of the Foxworth home invasion was the same person who raped and killed Holly Staker in Waukegan in 2000, and that was not Williford.

Importantly, while Williford maintains that the DNA "hit" from the prior testing showing implicate the facts of the Holly Staker case and show his innocence, *see* Exhibit 8, the probative value of these new genealogical results do not depend upon an inference to be drawn from being the same person committing multiple heinous crimes. Four reports from the private lab have currently been provided to Williford.[6] Those reports also led to Waukegan conducting an extensive investigation trying to locate and identify this person. Fox 32 News, *Holly Staker Murder: New Lead Gives Waukegan Police Hope In Solving Cold Case* (Aug. 17. 2022). Williford does not know the extent of the investigation undertaken by the Waukegan Police Department.[7] Nor will Williford publicly reveal the contents of these genealogical testing reports. Suffice to say that: (1) the perpetrator Conners identified was originally described as being "yellow skinned," R.662, 684, 686; (2) Williford's counsel argued this yellow-skinned person could not have been Williford, *e.g.*, R.1189, 1191, 1194, 1196, & 1207; and (3) the genealogical reports provide substantial new support to this misidentification argument.

The evidence is a game changer on its own. Williford is entitled to relief, as he is innocent. Anew trial ordered given this evidence. To be sure, given the focus on what happened at trial, there should be no question that the new evidence paints the jury verdict in an entirely different light because this very issue—the skin color and potential race of the perpetrator—was repeatedly raised as an issue at trial. In the original proceedings, Williford's counsel only had general arguments to make his case. Now, Williford has decisive DNA evidence from the murder weapon in a sample mixed with the victim's blood.

[6] These Reports were previously provided to the Court confidentially, via email. The reports are subject to protective order and Williford can separately provide them if convenient for the Court.

[7] The new reports and absence of information about the ongoing investigation led to additional litigation between Williford and the State. Among other things, Williford sought to compel further discovery from the State about its investigation into the identity of the person whose DNA was found at both the Foxworth home invasion and the Staker homicide. By Order of May 17, 2023, those requests were denied. As that litigation made clear, and noted in this Court's order, the instant motion was contemplated at the conclusion of the discovery-related litigation about the private genealogical reports. *Id.* at 3.

In short, the DNA evidence—the new NIRCL testing, Dr. Reich's Analysis, and the private genealogical reports—provide powerful evidence illustrating not only that Williford is innocent but demonstrate that the "identification evidence" in this case was mistaken and unreliable.

<u>Fourth</u>, there is new evidence that further undermines the identification and corroborates Williford's <u>Brady</u> claim. A little background: Officer Warner was the officer who first presented Williford's name to other investigators. Ex. 10 (Warner CI Reports). In so doing, Officer Warner claimed he had a telephone call with a "confidential informant" (Scott Henderson) on January **26**, 2000 (two days after the home invasion), that revealed only the <u>first name</u> of the suspect. *Id.; see* R.741. There was a dispute at trial about this alleged informant, R. 742. The objective evidence shows that Warner was lying about when he decided to treat Williford as a suspect—Warner had requested Williford's information in the LEADS system the day before, on January **25**, 2000. Ex. 11 (Warner LEADs receipt).

By 2015, Officer Warner admitted part of his misdeeds, and admitted that he did not have a phone call with an informant but had a conversation with Henderson *in person*, and with another woman present. Ex. 12 (Warner Task Force Report). That woman was not disclosed in any report or as having been present for the conversation with Warner and Henderson. And, it turns out that woman was having an extramarital sexual relationship with Warner, but still refuses to publicly admit these facts. Exhibit 13 (Affidavit of Jennifer Blagg, 2024). We now know, however, some of the reasons why Warner had motive to falsely implicate Williford—Williford was also sleeping with Warner's paramour. Exhibit 14 (Williford Affidavit).[8] Williford did not learn about this potential motive for Warner targeting him until after the trial. *Id.*

[8] Williford's affidavit will be provided to the Court subsequently.

These revelations—which should have been disclosed—are important Brady issues, not only because they would have impeached the testimony of several state witnesses, but because they would have revealed a new witness Williford could have considered calling at trial. Perhaps more important, these corrupt motives by investigators cast further doubt on the reasons why the officers conducted an unfair lineup procedure and only targeted one suspect throughout the case.

DISCUSSION

I. Legal Standards

While Illinois law typically contemplates one postconviction petition, fundamental fairness requires the bar against successive petitions be lifted where a petitioner raises a claim of actual innocence or by meeting the cause and prejudice standard from 725 ILCS 5/122-1(f). Put differently, a showing of cause-and-prejudice is required for constitutional claims but not an actual innocence claim. People v. Coleman, 2013 IL 113307, ¶ 91; People v. Pitsonbarger, 205 Ill.2d 444, 460 (2002).

At this juncture, where the State is not involved, the Court's threshold inquiry is limited to conducting a "*prima faciae* showing of cause and prejudice"; *i.e.*, whether the motion "adequately alleges facts demonstrating cause and prejudice." People v. Bailey, 2017 IL 121450, ¶24.

Put differently, the Court's threshold screening inquiry only determines whether the petition will advance to the normal three-stage process for evaluating post-conviction petitions, at which point the State can weigh in and Answer, seek dismissal, concede the petition, etc. See id. ¶26.

II. Williford Should be Granted Leave to File a Successive Post-Conviction Petition

Williford seeks to file a successive post-conviction petition raising the following claims:

(1) actual innocence,

(2) due process/fair trial concerning the admission of unreliable identification evidence,

(3) ineffective assistance of counsel, and

(4) due process/fair trial concerning the suppression of material information.

A. All of the New Information and Evidence Call into Question the Identification Evidence and Would Enable the Jury To Easily Conclude Williford Was Misidentified

In several ways, the new evidence—social science concerning memory and cognition, DNA, *Brady* material—shows how the single most important piece of evidence in the case was fatally flawed and put Conner's "identification" in a new light. Though counsel and police knew and should have known about the misidentification risk at the time from the procedures they used, the new social science conclusively proves that Conners' identification was unreliable. Once Conners had been shown a single picture of Williford and seen that picture again in September of 2022, her memory was irretrievably contaminated. And, Conners' behavior during the 2003 photo array bespeaks unreliability.

Though the DNA evidence stands alone and warrants relief independently, the DNA evidence provides further objective, forensic proof that could easily cause jurors to have reasonable doubt about Conners' purported identification. Indeed, the one question the jury asked in deliberations was about the picture of Williford. As a result, in a case with DNA all over the "murder weapons," and with sophisticated science, it is now more significant than it was at the original trial that Williford is excluded from contributing to this evidence. Meanwhile, two other people (Unidentified Male #1 and "Stain D") have their DNA mixed with Foxworth's blood. What information we have about Unidentified Male #1 also casts Conners description of the perpetrator in an entirely new light. What might have seemed odd to the jury in the original proceedings—questioning about a "yellow skinned man"—takes on an entirely new character in light of the new genealogical testing reports.

B. Williford Raises a Compelling Actual Innocence Claim

There is no bar to raising a successive actual innocence claim. So, advancing that should not be controversial. Williford should be given leave to file an actual innocence claim based upon the extensive DNA testing results in this case, the advancements in social science, witness statements from the Task Force investigation in 2015-2016, other evidence, and the intervening Illinois Supreme Court decision in Robinson that overruled the *dicta* in cases upon which this Court previously relied in denying Williford's petition.

C. Williford's Constitutional Claims Satisfy the Cause-And-Prejudice Standard

Williford's constitutional claims should be permitted as well.

1. Summary of Constitutional Claims

In different ways, each claim involves the misidentification of Williford by Conners at trial and its unreliability. The unreliability of the identification evidence supports (1) claims concerning how Conners's misidentification was generated through three identification procedures (*i.e.*, the due process issue under Manson and Neil), (2) the failure of defense counsel to seek suppression of this evidence (*i.e.*, the ineffective assistance issue), and (3) the reasons why Williford was a suspect in the first place (*i.e.*, the Brady issue). Brief elaboration follows.[9]

[9] Given the extensive discussion of the nature of the genealogical DNA reports necessary to present Williford's successive post-conviction petition, which will have to be filed under seal or heavily redacted, the successive petition is not appended to this motion.

First, from a due process/fair trial lens, the federal and state constitutions both prohibit unreliable evidence from being admitted against a criminal defendant. Manson v. Brathwaite, 432 U.S. 98, 116 (1977); Neil v. Biggers, 409 U.S. 188, 198 (1972); People v. Miller, 254 Ill. App.3d 997, 1003 (1993). The new scientific consensus confirms that Conners' misidentification of Williford at trial was irretrievably contaminated by the prior identification procedures; Conners' own hesitation in picking Williford from the third procedure is separate evidence on unreliability; and that there is simply no way to "purge" the prior, unfair procedures from Conners' testimony, regardless of how confident she appeared at trial.

Second, counsel is ineffective where a suppression motion should have been granted and the outcome of the trial probably would have been different as a result. People v. Givens, 384 Ill. App. 3d 101, 108 (2008). The additional evidence illustrating that Conners' identification was unreliable—both from the social science and the DNA evidence—confirms that Williford was prejudiced by his counsel's failure to seek suppression of this evidence. In addition, part of this Court's prior rejection of this claim turned on the fact that the 2002 identification procedure, memorialized in a police report and undisputed by the State, is not in the trial record. But, there is no question this event occurred.

Third, the Brady issue in the case concerns police investigators falsely writing in their reports and then testifying at trial that they learned Williford's name from a "Confidential informant" via a telephone call but did not have complete background information, including a photograph, of Williford until May of 2000. But, we now know, that the claims in these documents about an informant were in person; that they were in the presence of a third person who one of the officers (Warner) was having extramarital sex with; and that this same officer was upset at Williford for also having sex with the same woman whom he was having extramarital sex with. Williford did not learn of this fact until after his criminal trial. Officer Warner did not admit the true nature of the "confidential informant" meeting until he was contacted by the Task Force in 2015. Ex. 12. In addition, the paramour has confirmed in conversations with Williford's counsel that these events took place, though she is unwilling to voluntarily put them into writing. Ex 13.

The Brady material matters. It is often difficult for juries to think that police would have a reason to pursue a suspect who is innocent. But, the type of previously-suppressed evidence here—revenge for having sex with someone's sexual partner—is the type of classic tale that would make sense to a jury and have impeached the officers dearly. In the end, the bias of the investigators synchs up with the flawed show-up procedure in this case and, if known at trial, would have given Williford's counsel the opportunity to cross-examine the officers not only about the identification procedures but also their bias against Williford reflected in their false reports about the "confidential informant."

2. A *Prima Facie* Showing of Cause and Prejudice Standard Has Been Amply Made

Each of these claims satisfy the cause-and prejudice test, and this motion easily exceeds the *prima facia* showing required at this juncture. Cause is defined as an objective factor, external to the defense, that impeded the defendant's effort to raise the claim in an earlier proceeding. Pitsonbarger, 205 Ill.2d at 460. And, though what "constitutes cause will necessarily depend on the unique circumstances of each case," the Illinois Supreme Court "has observed 'that a showing that the factual or legal basis for a claim was not reasonably available to counsel' will constitute cause." People v. Blalock, 2022 IL 126682, ¶ 39 (quoting Pitsonbarger, 205 Ill. 2d at 460). Prejudice is defined as an error so serious that it affected the entire trial to the extent that the resulting conviction violates due process. Pitsonbarger, 205 Ill. 2d at 464.

For cause, the 2020 DNA testing, Dr. Reich's review of that testing, and the genealogical testing results did not exist in 2015 and the genealogical reports were only recently produced to Williford. Likewise, substantive developments in the science of eyewitness identification—including a path-breaking "White Paper" published in 2020 concerning multiple identification tests—constitute factors external to the defense that impeded Williford from raising this evidence. Quite simply, the 2017, 2020, and 2021 studies—and scientific consensus reflected therein—did not exist at the time Williford filed his prior petition in 2015. Nor Could have Williford discovered the *Brady* evidence, withheld from him at trial and secreted for over a decade and a half, until the witnesses were willing to admit the truth. This certainly satisfies the cause standard. See, e.g., People v. Wrice, 406 Ill. App. 3d 43, 52, (2010) (cause established when the defendant raised for the first time the argument new report corroborated his claims).

For prejudice, the misidentification was the heart of the State's case at trial. As the Court of appeals recognized, and unchanged by the stage of the proceedings, "absent [Conners'] testimony identifying [Williford] as one of the perpetrators, there is a reasonable probability that defendant would have been acquitted" because, "[q]uite simply, her testimony and her identification of [Williford] made up the majority of the State's case." Wiulliford II, at 13-14. Evidence showing this was unreliable or should have never been admitted is the quintessential definition of prejudice—it would unravel the cornerstone of this wrongful conviction.

In addition, despite the misidentification of a sincere (but mistaken) witness, the jury still exhibited some hesitation in voting to convict, and specifically inquired about the black-and-white picture they knew had been shown to Conners one time (even though, in truth, it had been shown to her twice). R.1242-43. Evidence showing Conners' identification was unreliable should have led to its suppression, but counsel failed to file a motion in that regard. Regardless, the DNA evidence, the scientific evidence concerning the unreliability of this particular claimed identification, and the evidence about the officers' motives to implicate Williford (and their lies about that fact) would have enabled the jury to see that the identification was unreliable and helped the jury understand that Conners' memory was contaminated by police officers who had been lying about their investigation from day one. There can be no reasonable dispute that evidence conclusively demonstrating Conners' identification was erroneous would probably change the result upon retrial. That is sufficient to prejudice. See e.g., People v. Brandon, 2021 IL App (1st) 172411 ¶¶71-77 (prejudice for purposes of filing successive petition shown by evidence that, if credited, would undermine the primary evidence against the defendant at trial).

CONCLUSION

Petitioner Marvin Williford respectfully requests that this Court grant him leave to amend an successive petition for post-conviction relief, asserting claims of (1) actual innocence, (2) violation of his right to a fair trial/due process related to the admission of unreliable evidence, (3) violation of his right to a fair trial due to ineffective assistance of counsel, and a violation of his right to due process under Brady/Giglio due to the suppression of material information that was both exculpatory and impeaching of prosecution witnesses.

Respectfully submitted,

Counsel for Petitioner

David B. Owens
THE EXONERATION PROJECT
311 N. Aberdeen Street, 3FL
Chicago, IL 60607
Telephone: (312) 789-4955
Attorney ID: 44407

CERTIFICATE OF SERVICE

I, David B. Owens, an attorney, hereby certify that on January 25, 2024, I caused to be served the foregoing Petitioner Marvin Williford's Motion for Leave to File Successive Amended Post-Conviction Petition on counsel of record for the People of the State of Illinois, as listed below.

One of Petitioner's Attorneys

Persons served:

Eric Rinehart
Barbara Buhai
18 N. County Street, 3rd FL
Waukegan IL 60085

EXHIBIT 1

North Chicago Police Department

Description INTERVIEW W/DENA CONNERS/CMDR.HOLDERBAUM **SUPPLEMENTAL REPORT**

Date Entered 2:52:55 PM, 09/20/2002 User HILL Case#: 2K-002716

Narrative

NORTH CHICAGO POLICE DEPARTMENT
CRIMINAL INVESTIGATIONS DIVISION
CASE#: 2K-002716
DATE: 9-19-02 @ 4:57PM
CMDR.HOLDERBAUM #304

R/CMDR WENT TO 611 S.ELMWOOD AND SPOKE TO DENA CONNERS, F/B, 7-23-58 OF THE SAME ADDRESS, TX: (847)623-6776.
THIS MEETING WAS FOR THE PURPOSE OF MS.CONNERS REVIEWING A STATEMENT SHE GAVE TO DET.L.WADE ON 1-22-2000.
MS.CONNERS REVIEWED HER STATEMENT AND REPLIED THE INFORMATION CONTAINED IN THE REPORT WAS CORRECT.
SHE ALSO STATED THAT SHE IDENTIFIED A PHOTOGRAPH OF THE OFFENDER WHO SET HER BOYFRIEND DELWIN FOXWORTH ON FIRE.
SHE WAS ADVISED THAT I HAD SIX PHOTOGRAPHS THAT I WOULD LIKE HER TO LOOK AT. I ASKED HER IF SHE COULD IDENTIFY ANYONE IN THE PHOTOS AS THE PERSON WHO SET FOXWORTH ON FIRE?
SHE THUMBED THROUGH THE PHOTOS AND PICKED OUT THE PHOTO OF MARVIN T.WILLIFORD AND STATED HE IS THE ONE WHO SET FOXWORTH ON FIRE. R/CMDR HAD OBTAINED THE PHOTOS FROM THE CICERO POLICE DEPARTMENT.
THE PHOTO WAS A MUG SHOT OF WILLFOLD DATING BACK TO 1990. SHE STATED THE OFFENDER WAS HEAVIER BUILT AT THE TIME FOXWORTH WAS SET ON FIRE. HOWEVER, SHE IS POSITIVE HE IS THE SAME PERSON.

THE PHOTOS SHOWN TO HER WERE OF;
ANTHONY JONES
DARRYL JORDAN
ALLEN JONES
MARVIN T.WILLIFORD
SETH HONES
LARRY JONES

Officer Signature (X) ______________________________

92S109

EXHIBIT 2

North Chicago Police Department

Description ADD - ON (PHOTO LINE-UP) CMDR.HOLDERBAUM#304

SUPPLEMENTAL REPORT

Date Entered 10:53:38 AM, 02/17/2003User HILL

Case#: 2K-002716

Narrative

NORTH CHICAGO POLICE DEPARTMENT
CRIMINAL INVESTIGATIONS DIVISION
DATE: 2-13-03 @ 8:15AM
HOMICIDE INVESTIGATION
SVC#: 2K-002716 (1-22-00)
CMDR.HOLDERBAUM #304

ON THE ABOVE DATE AND TIME, R/CMDR HOLDERBAUM MET WITH DELIA CONNERS, F/B OF [redacted] AT THE NORTH CHICAGO POLICE DEPT INVESTIGATIONS DIVISION.

CONNERS WAS WITH DELWIN FOXWORTH THE NIGHT HE WAS SET ON FIRE AND WITNESSED THE INCIDENT. CONNERS WAS ADVISED SHE WOULD BE SHOWN A PHOTO LINE UP CONSISTING OF SIX PHOTOGRAPHS OF MALE BLACK INDIVIDUALS.
SHE WAS ASKED TO VIEW THE LINE UP AND IF POSSIBLE IDENTIFY THE PERSON SHE KNOWS AS "T" WHO WAS THE PERSON WHO DOUSED DELWIN WITH GASOLINE AND SET HIM ON FIRE.

CONNERS VIEWED THE LINE UP FOR SEVERAL SECONDS. SHE THEN SAID IT LOOKS LIKE "T", IS EITHER NUMBER #4 OR #5. SHE SAID FOUR LOOKS LIKE HIM HOWEVER THE FACE IS TO THICK.
SHE LOOKED AT NUMBER #5 AND STATED THAT #5 IS THE PERSON WHO SET DELWIN ON FIRE.

SHE STATED SHE WAS POSITIVE. CONNERS WAS THANKED FOR ASSISTANCE AND CLEARED THE POLICE DEPT.

Officer Signature (X) ______________________________

9SS105

EXHIBIT 3

American Psychological Association · American Psychology-Law Society · Law and Human Behavior

© 2020 American Psychological Association
ISSN: 0147-7307

2020, Vol. 44, No. 1, 3–36
http://dx.doi.org/10.1037/lhb0000359

Policy and Procedure Recommendations for the Collection and Preservation of Eyewitness Identification Evidence

Gary L. Wells
Iowa State University

Margaret Bull Kovera
John Jay College and the Graduate Center, City University of New York

Amy Bradfield Douglass
Bates College

Neil Brewer
Flinders University

Christian A. Meissner
Iowa State University

John T. Wixted
University of California, San Diego

Objective: The Executive Committee of the American Psychology-Law Society (Division 41 of the American Psychological Association) appointed a subcommittee to update the influential 1998 scientific review paper on guidelines for eyewitness identification procedures. ***Method:*** This was a collaborative effort by six senior eyewitness researchers, who all participated in the writing process. Feedback from members of AP-LS and the legal communities was solicited over an 18-month period. ***Results:*** The results yielded nine recommendations for planning, designing, and conducting eyewitness identification procedures. Four of the recommendations were from the 1998 article and concerned the selection of lineup fillers, prelineup instructions to witnesses, the use of double-blind procedures, and collection of a confidence statement. The additional five recommendations concern the need for law enforcement to conduct a prelineup interview of the witness, the need for evidence-based suspicion before conducting an identification procedure, video-recording of the entire procedure, avoiding repeated identification attempts with the same witness and same suspect, and avoiding the use of showups when possible and improving how showups are conducted when they are necessary. ***Conclusions:*** The reliability and integrity of eyewitness identification evidence is highly dependent on the procedures used by law enforcement for collecting and preserving the eyewitness evidence. These nine recommendations can advance the reliability and integrity of the evidence.

Public Significance Statement

Mistaken eyewitness identification is a primary contributor to criminal convictions of the innocent. Pristine procedures for collecting and documenting eyewitness identification evidence can help prevent these mistakes. This scientific review paper makes nine system variable recommendations concerning eyewitness identification procedures that should be implemented by crime investigators in eyewitness identification cases.

Keywords: eyewitness identification, lineups, showups, identification procedures, eyewitness recommendations

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

Gary L. Wells, Department of Psychology, Iowa State University; Margaret Bull Kovera, Department of Psychology, John Jay College and the Graduate Center, City University of New York; Amy Bradfield Douglass, Department of Psychology, Bates College; Neil Brewer, College of Education, Psychology and Social Work, Flinders University; Christian A. Meissner, Department of Psychology, Iowa State University; John T. Wixted, Department of Psychology, University of California, San Diego.

This is an official statement of the American Psychology-Law Society, Division 41 of the American Psychological Association, and does not represent the position of the American Psychological Association or any of its other Divisions or subunits.

The authors thank the following individuals for their comments and suggestions on earlier versions of this article: Alexis Agathodeous, Assistant District Attorney David Angel, Dominick Atkinson, Jen Beaudry, Shari Berkowitz, Iris Blandón-Gitlin, Chief William Brooks, Brian Cahill, Steve Charman, Brian Cutler, Jen Dysart, Mitch Eisen, Melanie Fessinger, Ron Fisher, Nancy Franklin, Valerie Hans, Kate Houston, Jennifer Jones, Detective Matt Jones, Saul Kassin, Elizabeth Loftus, Jamal Mansour, Karen Newirth, Steve Penrod, Dan Reisberg, Andrew Smith, Siegfried Sporer, Nancy Steblay, Jennifer Teitcher, and Adele Quigley-McBride.

Correspondence concerning this article should be addressed to Gary L. Wells, Department of Psychology, Iowa State University, West 112 Lagomarcino, Ames, IA 50011. E-mail: glwells@iastate.edu

The identification of criminal suspects by eyewitnesses continues to be a staple form of evidence used by the justice system to help establish the identity of crime culprits. Over the last few decades, however, serious concerns have been raised about the potential unreliability of eyewitness identification evidence. There have been two primary forces that have helped to shape this concern about the reliability of eyewitness identifications. First, psychological scientists began developing programmatic laboratory-based experimental research starting in the mid to late 1970s that focused on eyewitness identification. Early in this research, it became apparent that certain variables under control of the criminal justice system can dramatically inflate the likelihood of a mistaken identification; these variables were called *system variables* (Wells, 1978). An instruction given to eyewitnesses prior to viewing a lineup, for example, is a system variable because it influences the reliability of the eyewitness's identification and is under the control of the legal system's policies and procedures for administering lineups (Malpass & Devine, 1981). By the mid-1990s, psychological scientists had published hundreds of laboratory-based experiments in peer-reviewed journals showing that mistaken identification rates can be very high under certain conditions and had identified some of the more problematic sets of conditions that can lead to such errors in real-world circumstances (e.g., Cutler & Penrod, 1995).

The second major force propelling a strong awareness of the potential fallibilities of eyewitness identification was the application of forensic DNA testing to claims of innocence. Although forensic DNA testing was conceived primarily as a tool to incriminate the guilty, the exonerating powers of forensic DNA testing quickly became evident. Starting in the 1990s, DNA was used to test claims of innocence in selected postconviction cases and a cascade of exonerations of innocent people began to unfold. In a report commissioned by the U.S. Department of Justice examining the first 30 exonerations, it was quite clear that the vast majority of these cases involved mistaken eyewitness identifications (Connors, Lundregan, Miller, & McEwan, 1996). Although legal scholars had described a number of cases of innocent people being convicted in the United States based on mistaken identification before the development of DNA testing (e.g., Borchard, 1932; Brandon & Davies, 1973; Frank & Frank, 1957; Huff, Rattner, & Sagarin, 1986), the DNA exoneration cases quickly outnumbered previously known cases of convictions of the innocent. Moreover, the pre-DNA cases tended to be less definitive about whether the person was, in fact, innocent, often just indicating "probable innocence" or "legal innocence" rather than the more definitive "actual innocence" characterization that accompanied DNA-based exonerations (Garrett, 2011; Scheck, Neufeld, & Dwyer, 2000; West & Meterko, 2017).

The development of a scientific literature on eyewitness identification and the use of forensic DNA testing to uncover mistaken eyewitness identifications were a powerful combination. In 1996, the Executive Committee of the American Psychology-Law Society (AP-LS, Division 41 of the American Psychological Association), the primary scholarly organization for eyewitness identification researchers, appointed a committee to draft a scientific review paper on recommendations for how to collect eyewitness identification evidence. The result of that review, vetted by the membership of AP-LS and subjected to peer review, was published in *Law and Human Behavior* (Wells et al., 1998). That article, commonly referred to as the "white article on lineups," is the forerunner to the current article. In effect, the current article is an update of the original scientific review paper on lineups and, like the original scientific review paper, also represents the official position of the AP-LS on these issues. The 1998 scientific review paper was the first set of science-based recommendations regarding how to conduct lineups that was endorsed by a scientific society. In addition, the 1998 scientific review paper played an important role in subsequent developments. For example, it was the model on which the U.S. Department of Justice made its recommendations in its 1999 guide for law enforcement on collecting and preserving eyewitness evidence. The NIJ Guide was mailed to all of the more than 17,000 law enforcement agencies in the U.S. (Technical Working Group for Eyewitness Evidence, 1999). Moreover, the 1998 article was a model on which the New Jersey Department of Justice created the first statewide guidelines on eyewitness identification that law enforcement were required to follow. It was also the first peer-reviewed publication to include a description of the unfolding DNA exoneration cases (40 at that time) and what these DNA cases might tell us about eyewitness identification evidence. The 1998 scientific review paper has been used as supporting written material in countless seminars and workshops of lawyers, police, prosecutors, and judges, as well as in police training. Finally, the 1998 scientific review paper has been relied on by state and federal courts across the U.S. (e.g., *New Jersey v. Henderson*, 2011; *State v. Lawson*, 2012; *Young v. Conway*, 2013).

The Need for an Updated Scientific Review Paper on Lineups

Today, our understanding of eyewitness identification has matured well beyond where it was when the previous scientific review paper was published. Clearly, experimental laboratory studies have grown immensely in number and breadth over the last 20 years. However, the general methodology of laboratory eyewitness identification studies has largely remained the same. People are exposed to a simulated crime, sometimes live, sometimes a video, and the researchers know the true identity of the culprit. These participant-witnesses are then shown an identification procedure, typically a photo lineup, in which the culprit's photo is embedded among filler photos or the culprit's photo is absent from the array and replaced with the photo of an innocent person. Using this basic paradigm, researchers then systematically manipulate variables, such as the view the witness had of the culprit, the similarity of the fillers to the culprit, the instructions given to the eyewitness prior to viewing, suggestive behaviors of the lineup administrator, and so on, to see how those variables affect the responses of the eyewitnesses. This experimental laboratory methodology has a number of strengths that arise from the fact that "ground truth" is known (i.e., the researchers know which person is the actual "culprit") and from the use of random assignment to conditions that permit inferences about the cause of any effects observed in responses.

One type of data that was largely unavailable at the time of the 1998 review comes from published field studies of police lineups. By field study we mean either an archival or prospective examination of the results of lineups conducted by police investigators in actual cases. To be included as a field study in our analysis, the

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

study had to report how often the eyewitnesses identified the suspect, identified a filler, or made no identification and the study had to be published in a peer-reviewed journal. Only one field study had been published in a peer-reviewed journal at the time of the 1998 review but now there are 11 peer-reviewed published field studies. In these field studies of lineup outcomes, the lineup contains "fillers" who are known to be innocent of the crime (such as people who were in prison at the time of the crime). If the eyewitness identifies the suspect, it might or might not be an accurate identification because ground truth is not known with certainty in actual cases. But when an eyewitness identifies a filler, it is clearly a mistaken identification because these fillers are known to be innocent. A mistaken identification of a filler will not result in charges against that filler, but a filler identification is a mistaken identification nevertheless. These field studies have clear limits, in large part because of the difficulty of establishing ground truth for identifications of suspects (i.e., not knowing whether the suspect is innocent or guilty, see Horry, Halford, Brewer, Milne, & Bull, 2014). However, these field study data help counter some of the criticisms of laboratory-based experiments—in particular, the criticism that participants in laboratory experiments make a lot of mistaken identifications only because the witnessed events are not real and the consequences of mistaken identification are trivial (Mecklenburg, Bailey, & Larson, 2008). The argument is that real witnesses to serious crimes would not be so careless. Hence, perhaps laboratory studies vastly overestimate the eyewitness misidentification problem. The results of peer-reviewed field studies, however, show otherwise.

Based on these field studies, we can now estimate how often actual eyewitnesses in serious crime cases mistakenly identify a filler from a lineup. These 11 peer-reviewed published studies collected data from a total of 6,734 lineups. These field studies are from highly varied jurisdictions (e.g., California, Arizona, Texas, London, England) and a summary of these data is shown in Table 1. For current purposes, two statistics of note from Table 1 speak to the question of whether actual witnesses to serious crimes are too cautious to make mistaken identifications at rates like those observed in lab experiments. First, nearly one of every four witnesses (23.7%) who was shown a lineup selected an innocent filler. Second, among those who made an identification (35.5% made no identification), over one third (36.8%) identified a known-innocent filler. A summary of 94 laboratory eyewitness identification studies showed that filler identification rates averaged 21.2% when the culprit was present and 34.6% when the culprit was absent (Clark, Howell, & Davey, 2008). Averaging these two filler identification rates from lab studies yields 27.9% filler identifications in the average laboratory study versus 23.7% filler identifications found in actual cases. These field study data, which were not available at the time of the 1998 scientific review paper, suggest that experimental laboratory studies are not producing highly inflated rates of mistaken identification compared with what happens with actual eyewitnesses to serious crimes. Of course, lab experiments and actual eyewitness identification cases differ in many ways. For example, actual eyewitness cases involve longer retention intervals, more violence, and more guns than do lab studies (see Flowe, Carline, & Karoglu, 2018). Nevertheless, lab studies are particularly valuable for isolating cause-effect relations among variables, which is a feature that tends to elude field studies for a variety of reasons (see Horry et al., 2014).

Table 1
Summary Statistics on 11 Published Field Studies of Eyewitness Identification

Authors	Number of possible IDs	ID of suspect	IDs of filler	No ID	Suspect %	Filler %	No ID%	% Making an ID	Suspect ID rate among all IDs	Filler ID rate among all IDs
Behrman and Davey (2001)	58	29	14	15	50.0%	24.1%	25.9%	74.1%	67.4%	32.6%
Behrman and Richards (2005)	461	238	68	155	51.6%	14.8%	33.6%	66.4%	77.8%	22.2%
Horry, Halford, Brewer, Milne, and Bull (2014)	833	382	149	302	45.9%	17.9%	36.3%	63.7%	71.9%	28.1%
Horry, Memon, Wright, and Milne (2012)	1,039	406	273	360	39.1%	26.3%	34.6%	65.4%	59.8%	40.2%
Klobuchar, Steblay, and Caligiuri (2006)	178	63	20	95	35.4%	11.2%	53.4%	46.6%	75.9%	24.1%
Memon, Havard, Clifford, Gabbert, and Watt (2011)	1,044	456	437	151	43.7%	41.9%	14.5%	85.5%	51.1%	48.9%
Valentine, Pickering, and Darling (2003)	584	237	121	226	40.6%	20.7%	38.7%	61.3%	66.2%	33.8%
Wells, Steblay, and Dysart (2015)	494	132	75	287	26.7%	15.2%	58.1%	41.9%	63.8%	36.2%
Wixted, Mickes, Dunn, Clark, and Wells (2016)	348	114	104	130	32.8%	29.9%	37.4%	62.6%	52.3%	47.7%
Wright and McDaid (1996)	1,561	611	310	640	39.1%	19.9%	41.0%	59.0%	66.3%	33.7%
Wright and Skagerberg (2007)	134	78	28	28	58.2%	20.9%	20.9%	79.1%	73.6%	26.4%
Overall sum	6,734	2,746	1,599	2,389						
Weighted means					40.8%	23.7%	35.5%	64.5%	63.2%	36.8%

Note. Some studies reported data that included identifications by witnesses who knew the culprit (prior familiarity) and those data are excluded from Table 1.

In addition to a large growth in the number of peer-reviewed laboratory-based experiments, as well as the addition of field studies of lineups in actual cases, DNA exoneration cases involving mistaken eyewitness identification have continued to accumulate since the 1998 review paper. As of this writing, DNA has established that at least 365 people in the U.S. were convicted and imprisoned (some on death row) for crimes they did not commit (Innocence Project, 2019). More than 70% of these DNA exoneration cases involved mistaken eyewitness identification. Although the number of DNA exonerations represent only a small fraction of convictions, it is essential to note that these exonerees were the lucky few for whom DNA-rich trace evidence for the crime existed, was collected and preserved properly, and was tested. Contrary to public perceptions based on TV programs, culprits leave behind DNA-rich trace evidence in only a small fraction of cases. The largest category of convictions based on eyewitness identification evidence is robberies, and culprits of robbery almost never leave behind DNA-rich trace evidence that could exculpate a mistakenly identified person. For cases in which DNA did exist (primarily restricted to sexual assault cases), no one anticipated that forensic DNA testing would later develop and so for many cases prior to the advent of forensic DNA testing, the trace evidence was not collected, was not preserved properly (allowing it to deteriorate), was lost, or was destroyed by the time forensic DNA testing came along. Even today, individuals who claim innocence may not be permitted by statute or discretionary decision (by a court or a prosecutor) to access and test available DNA evidence. Hence, DNA testing could only reverse a very small fraction of possible mistaken identifications. Indeed, eyewitness identification evidence is still heavily relied upon today because DNA and other forms of definitive evidence remain extremely rare. Improving the reliability of eyewitness identification evidence therefore remains an important priority in preventing miscarriages of justice.

Focus of the Current Article

The current article, like the 1998 scientific review paper, is not a broad review of all issues in eyewitness identification. It is, instead, a focused examination of *system variables* in eyewitness identification: factors that relate to the reliability of eyewitness identifications over which the justice system has (or can have) control. Hence, despite the fact that there are many powerful variables that affect the reliability of eyewitness identification evidence, which are called *estimator variables* (e.g., same vs. cross-race identifications, stress during the witnessing of a crime, quality of view), such variables are not under control of the justice system and are therefore not the focus of this scientific review paper.

In some ways, the definition of system variables that is used today has broadened from its original definition. Originally, system variables in eyewitness identification referred to variables that *influence the accuracy of eyewitness identifications* over which the justice system has control (Wells, 1978). Over time, however, the definition of system variables has broadened to include factors under the control of the justice system that *relate to* (as opposed to *influence*) the accuracy of eyewitness identifications. Eyewitness confidence, for example, does not itself influence the accuracy of an eyewitness identification; however, eyewitness confidence is related to the accuracy of eyewitness identification and it is easily contaminated by events that *are* under the control of the criminal justice system (such as feedback from the lineup administrator). Accordingly, securing a confidence statement at the time of identification using a double-blind lineup administrator is a system variable (see Wilford & Wells, 2013, for a more extended discussion of this broader view of system variables).

In the current article, other examples of this broader definition of system variables will become apparent. For example, one of the recommendations in the current scientific review paper is video-recording the entire identification procedure. Obviously, video-recording the identification procedure is not meant to increase the accuracy of eyewitness identifications. Instead, the purpose of the video-recording recommendation is to secure a record that might help to assess the quality of the identification and the procedure (Sporer, 1992, 1993). That is, a video can shed light on the likely accuracy of eyewitness identifications via creating a record of behavioral cues (such as decision time, spontaneous comments, and confidence cues) that are diagnostic of accuracy (Kaminski & Sporer, 2017). In addition, video-recordings of the identification procedure can help document that police followed recommended procedures and provide other potential benefits as discussed later in this article. Using the broader definition, video-recording qualifies as a system variable because it is under the control of the justice system to either video-record or not record, and the recording is relevant to the eyewitness accuracy problem.

A central issue in the development of recommendations on policies and procedures in eyewitness identification is how to decide which recommendations are the most important and what criteria should be used to decide whether to include a recommendation. For most of our recommendations, there is a solid and specific body of scientific evidence to support the recommendation and we review that scientific evidence. In some cases, however, the recommendation is based primarily on reasonably well-established understandings of human memory and social influence and our general understanding of problems that we have observed in actual cases. Consider, for example, our recommendation that the entire identification procedure be video-recorded. We believe that video-recording can have many benefits, including but not restricted to: moderating potential suggestive behaviors by the lineup administrator, establishing proof as to exactly what instructions were given to the eyewitness, recording information about how long it took the eyewitness to make an identification, and establishing both verbal and nonverbal records of the confidence expressed by the eyewitness. We believe that the arguments favoring this recommendation are compelling and elements of the recommendation are grounded in the science, such as the science showing that the verbal and nonverbal behaviors of the witness during the lineup are diagnostic of accurate versus mistaken identifications (e.g., Kaminski & Sporer, 2017). Hence, video-recording could be justified on that scientific ground alone. However, there has been little research on the issue of whether video-recording serves a prophylactic function, how the videos should be used, or on whom the camera should be focused. Hence, not every element of the benefits that we propose for video-recording of the identification procedure have been fully studied. Nevertheless, we believe that it would be irresponsible to not include a video-recording recommendation even if a subset of the benefits (e.g., its prophylactic function) has not yet been thor-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

oughly demonstrated. For each of the recommendations that we offer, we include justifications for the recommendation as well as a discussion of any concerns or caveats.

Live Versus Photo Lineups

Throughout this article, we use the term lineup to refer to both live lineups and photo lineups. We do not include separate sections on these two common procedures because we know of no evidence to indicate that the principles governing photo lineups and live lineups are different (see Fitzgerald, Price, & Valentine, 2018). In others words, every recommendation that we make about lineups applies equally to live and photographic displays.

Photo lineups are far more common than live lineups in most U.S. jurisdictions (Police Executive Research Forum, 2013). Photo lineups are sometimes called photo-spreads, photo-arrays, or photo-montages. Another common name for photo lineups in many U.S. law enforcement circles is "six packs," which refers to the most frequent size and arrangement of a photo lineup in which the lineup contains six photos arranged in two rows of three. Live lineups, sometimes called corporeal or physical lineups, are relatively rare (in research and in practice) compared with photo lineups, likely due to the greater difficulty, time, and expense of constructing and conducting live lineups relative to photographic lineups. But some jurisdictions, such as New York City, commonly conduct both live and photo lineups.

Even though the principles governing live versus photo lineups are thought to be the same, questions have been raised as to whether performance overall might be better for live lineups than for photo lineups. The *live-superiority hypothesis* predicts that the three-dimensional nature of a live showing (vs. two-dimensional photos) of the lineup members, along with having visual information about the full-bodies (not just faces) of the lineup members, would clearly render live lineups superior to photo lineups. However, a review of the scientific evidence comparing live with photo lineups (as well as video lineups) showed no support for the live-superiority hypothesis (Fitzgerald et al., 2018). In addition, there are a number of practical difficulties involved in organizing and administering live lineups, including the greater difficulty of finding appropriate fillers for live lineups, and the need to carefully orchestrate the timing and roles of various people at the lineup event. Some problematic elements are also difficult to control in live lineups, such as the appearance of nervousness on the part of the suspect, a nervousness not likely to be shared by fillers. Because of the apparent absence of any significant advantage in accuracy along with the greater practical difficulty of live lineups, "live lineups are rarely the best option in practice" (Fitzgerald et al., 2018, p. 307).

We take no position on live versus photo because preference for one over the other is likely to depend on the circumstances of the particular case. For example, if the witness described something distinctive about the body of the culprit or the culprit's gait, then a live lineup might be preferred. But the difficulty of constructing such a lineup with live fillers who match the witness's description of the culprit can be very challenging. Also, it is not uncommon for a suspect to be at-large, which precludes the use of a live lineup. In other cases, the stress induced when a victim-witness is asked to view their assailant live for purposes of identification might cause difficulties that could be avoided with a photo lineup. Some law enforcement agencies have used live lineups because there were reasons to believe that the witness might be able to identify the culprit's voice by having lineup members speak which, of course, is not possible while doing a photo lineup. Of course, doing a photo lineup does not preclude a later voice lineup with that suspect using only recorded voices and no visual information. In fact, some have argued that voice identification should be conducted separately from the visual lineup because the diagnosticity of the information obtained is greater if the witness can identify the voice and the face independently of each other (Pryke, Lindsay, Dysart, & DuPuis, 2004). Hence, conducting photo lineups does not preclude the identification of voices using a separate procedure.

Lineups as Distinguished From Showups

The 1998 scientific review paper dealt only with lineups, but there is another common identification procedure, called showups, that is included in this new scientific review paper. The basic distinction between lineups and showups is that lineups embed the suspect among known-innocent fillers whereas showups do not use fillers and instead simply present the suspect alone. There is no debate among eyewitness scientists about the fact that lineups produce better outcomes than do showups, whether the outcomes are measured in terms of diagnosticity ratios or measured using signal-detection based methods. There is some debate about the *process* by which the outcomes for lineups are superior to the outcomes for showups. For example, it has been suggested that the use of good lineup fillers can help the witness decide which facial features are relevant for making an identification decision (e.g., Wetmore et al., 2015; Wixted & Mickes, 2014). Others, however, note that lineups appear to be superior to showups only because a large share of mistaken identifications are siphoned off to fillers when lineups are used whereas *all* mistaken identifications land on the innocent suspect for showups because showups have no fillers (e.g., see Smith, Wells, Lindsay, & Penrod, 2017; Wells, Smith, & Smalarz, 2015). These two accounts of how lineups manage to produce better outcomes than showups are very different, but no eyewitness scientist contests the general observation that lineups with good fillers result in better applied outcomes than do showups.

If there is no question about the superiority of lineups over showups, why do we have recommendations about how showups should be conducted? Why not simply state that showups should never be conducted and that lineups should be conducted instead? There have been calls by some eyewitness scientists for the elimination of showups based on the clear evidence that showups are inferior to lineups (see Levi & Lindsay, 2001). But there are legal and policy reasons to permit showups under certain conditions even though a lineup would be more diagnostic. We review those reasons in the section that details the showups recommendation.

As a final note, it should be apparent that there should never be such a thing as a photographic showup. After all, the justification for a showup is that the individual has been detained on the street and there is a very limited time frame for conducting an identification procedure. If investigators are merely in possession of a photo of a suspect, there is no reasonable excuse for not taking the time to embed the photo among filler photos and conduct a proper photo lineup.

Overview of Recommendations

The 1998 scientific review paper had only four recommendations. The current scientific review paper endorses these same four recommendations but expands the number of recommendations to nine. The original four recommendations in the 1998 scientific review paper included: (a) the identification procedure should be administered using a *double-blind* procedure (i.e., the lineup should be administered by someone who does not know which person is the suspect and which persons are fillers); (b) *prelineup instructions* to the witness should emphasize that the culprit might or might not be in the lineup and that the lineup administrator does not know which person is suspected of being the culprit; (c) there should be only *one suspect per lineup and the suspect should not stand out* from the fillers based on the witness's description of the culprit or other factors that would draw attention to the suspect; and (d) a *confidence statement* should be secured from the witness at the time of identification and prior to any opportunity to get feedback about the identification decision. Although we include these four recommendations in our new set of nine, we have modified them in certain ways. For example, the double-blind recommendation now includes other means for accomplishing the goal of preventing influence from the lineup administrator that do not necessarily require a neutral administrator. The prelineup instructions include reworked language that is intended to make the instructions more effective. Finally, the securing of a confidence statement includes the recommendation of recording a confidence statement for both affirmative identification decisions and rejection decisions rather than only if the eyewitness makes an affirmative identification decision.

The original four recommendations in the 1998 scientific review paper were restricted almost exclusively to matters that occurred only during the lineup itself. Some of the five new recommendations, however, cover broader territory. For example, new recommendations concern matters that occur *before* the commencement of an identification procedure, including consideration of when it might be unwise to conduct an identification procedure, the problem of repeated identification procedures with the same witness and suspect, and the importance of conducting a proper interview of the witness prior to conducting the identification procedure. In addition, we make a recommendation concerning the appropriate use of showups.

The following is a brief description of each of the nine recommendations. The numeric order of the recommendations corresponds roughly to the temporal order in which police/administrators would likely encounter the matters covered by the recommendation (except for the last recommendation, which concerns showups).

1. ***Prelineup Interview Recommendation.*** Before conducting an identification procedure and as soon as practicable after the commission of the crime, an officer should interview witnesses to document their descriptions of the culprit, obtain their self-report of viewing conditions and attention during the crime, document any claims of prior familiarity with the culprit, instruct witnesses to not discuss the event with other cowitnesses, and warn the witnesses against attempting to identify the culprit on their own. The entire interview should be video-recorded.

2. ***Evidence-Based Suspicion Recommendation.*** There should be evidence-based grounds to suspect that an individual is guilty of the specific crime being investigated before including that individual in an identification procedure and that evidence should be documented in writing prior to the lineup.

3. ***Double-Blind (or Equivalent) Recommendation.*** Lineups should be conducted using a double-blind procedure (i.e., neither the administrator nor the witness should know who the suspect is in the lineup) or an equally effective method of preventing the lineup administrator from inadvertently influencing the witness.

4. ***Lineup Fillers Recommendation.*** There should be only one suspect per lineup and the lineup should contain at least five appropriate fillers who do not make the suspect stand out in the lineup based upon physical appearances or other contextual factors such as clothing or background.

5. ***Prelineup Instructions Recommendation.*** When inviting an eyewitness to attend a lineup procedure (photo lineup or live lineup), police should not inform the eyewitness of any information that the witness has not already provided and certainly should not suggest that the suspect who will be in the lineup has been arrested or that the culprit will be present in the identification procedure. The eyewitness should be instructed that (a) the lineup administrator does not know which person is the suspect and which persons are fillers; (b) the culprit might not be in the lineup at all, so the correct answer might be "not present" or "none of these"; (c) if they feel unable to make a decision they have the option of responding "don't know"; (d) after making a decision they will be asked to state how confident they are in that decision; and (e) the investigation will continue even if no identification is made.

6. ***Immediate Confidence Statement Recommendation.*** A confidence statement should be taken from witnesses as soon as an identification decision (either positive or negative) is made.

7. ***Video-Recording Recommendation.*** The entire identification procedure, including prelineup instructions and witness confidence statement, should be video-recorded.

8. ***Avoid Repeated Identifications Recommendation.*** Repeating an identification procedure with the same suspect and same eyewitness should be avoided regardless of whether the eyewitness identified the suspect in the initial identification procedure.

9. ***Showups Recommendation.*** Showups should be avoided whenever it is possible to conduct a lineup (e.g., if probable cause exists to arrest the person then a showup should not be conducted). Cases in which it is necessary to conduct a showup should use the procedural safeguards that are recommended for lineups, including the

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

elimination of suggestive cues, a warning that the detained person might not be the culprit, video-recording the procedure, and securing a confidence statement.

In the detailed treatments of the nine recommendations that follow, we placed a strong emphasis on communicating an underlying principle for each recommendation. The underlying principle for a recommendation is important because there can be times in which circumstances might require deviation from the literal specifics of the recommendation, but the deviation would still be able to conform to the underlying principle. For example, our recommendation for how to select fillers to use in a lineup is relatively specific but, in the end, it is more important that the underlying principle be achieved, namely that the fillers should be chosen in a way that would not make an innocent suspect stand out in the lineup. Hence, although we provide specific recommendations, it is more important to use procedures that reflect the principles behind the recommendations than to follow the specific recommendations.

Recommendations

Each of the nine recommendations begins with a statement of the recommendation. We then describe the rationale for the recommendation, including relevant data and the reasoning behind the recommendation. In addition, most of the recommendations have nuances or caveats, and some have practical concerns that are discussed.

Recommendation 1: Prelineup Interview

Before conducting an identification procedure and as soon as practicable after the commission of the crime, an officer should interview eyewitnesses to document their descriptions of the culprit, obtain their self-report of viewing conditions and attention during the crime, document any claims of prior familiarity with the culprit, instruct witnesses to not discuss the event with other cowitnesses, and warn the witnesses against attempting to identify the culprit on their own. The entire interview should be video-recorded.

In many cases there might be 911 (emergency call) recordings or initial witness statements by first responders that can prove useful to an investigation; however, this recommendation concerns a more extensive interview that would be conducted by an investigative officer. Recommendation 1 relates to the conduct of this interview with a witness or victim, during which time an investigator collects a statement relating to the person's memory for the event and the culprit(s). Collection of a detailed description of the culprit is a critical form of evidence that can facilitate investigators' attempts to locate a suspect (Brown, Lloyd-Jones, & Robinson, 2008; Kebbell & Milne, 1998). There is now substantial research on the most effective procedures for interviewing a witness or victim following an event (see Dando, Geiselman, MacLeod, & Griffiths, 2015; Fisher, Schreiber Compo, Rivard, & Hirn, 2014), as well as the harmful effects of suggestive or misleading interviewing procedures that should be avoided (see Brainerd & Reyna, 2005; Loftus, 2017; Newman & Garry, 2013). Specific interviewing procedures have also been developed for the collection of *person descriptions* (see Demarchi & Py, 2009; Gabbert & Brown, 2015; Meissner, Sporer, & Schooler, 2007; Satin & Fisher, 2019; Sporer, 1996). Although the current recommendations focus largely on the interviewing of adult witnesses, many of the same principles of memory apply to the interviewing of child witnesses (who are particularly susceptible to suggestion, see Ceci & Bruck, 1995). A robust literature is available for interested readers documenting the challenges of interviewing child witnesses (see Kask & Bull, 2009), including the development of effective, evidence-based protocols for interviewing children (see LaRooy et al., 2015; Sternberg, Lamb, Esplin, Orbach, & Hershkowitz, 2002).

The Contents and Accuracy of Person Descriptions

Obtaining an accurate and complete description of the culprit is important to furthering an investigation and ultimately can facilitate identification of the culprit. Archival studies suggest that witnesses tend to provide between seven and nine descriptors of a culprit on average, frequently including information about perceived height, weight, gender, ethnicity, and age (Fahsing, Ask, & Granhag, 2004; Granhag, Ask, Rebelius, Öhman, & MacGiolla, 2013; Sporer, 1992, 1996; van Koppen & Lochun, 1997; Yuille & Cutshall, 1986). Descriptions of the culprit's clothing, stature, and facial features are generally less frequent. When specific facial descriptors are provided, the majority refer to upper regions of the face, in particular the hair, eyes, and nose. Although estimates of height, weight, and age can be biased by the witness's own characteristics (e.g., individuals who are less than average height tend to underestimate height; see Flin & Shepherd, 1986), witnesses otherwise appear to provide an accurate, general impression of the culprit. Such descriptions, however, are often lacking in specific details (Douglass, Brewer, Semmler, Bustamante, & Hiley, 2013; Fahsing et al., 2004) that might prove useful for the construction and assessment of identification arrays (Corey, Malpass, & McQuiston, 1999), and it is therefore important that investigators use evidence-based procedures to enhance the quality of witnesses' accounts.

System Variables That Influence the Quality of Witness Accounts

Much like other memory phenomena, a host of factors can influence the accuracy and completeness of a witness's memory for the event and culprit (see Granhag, Ask, & MacGiolla, 2013; Meissner et al., 2007). Consistent with the general eyewitness literature, a distinction can be drawn between system and estimator variables (Wells, 1978). With respect to the former, the manner in which a witness is interviewed by an investigator can undermine the accuracy of a witness's statement. In particular, witnesses appear quite susceptible to the misinformation effect (see Berkowitz & Loftus, 2018; Loftus, 2017; Newman & Garry, 2013) in which leading or suggestive questioning from an investigator can distort memory reports, and witnesses can be induced to self-generate errors in their descriptions when forced or encouraged to provide a "complete" description of the event or culprit (Ackil & Zaragoza, 1998; Meissner, Brigham, & Kelley, 2001). In this respect, the use of facial feature checklists is not recommended, as they can subtly encourage "complete" responses that produce less accurate person descriptions (Wogalter, 1991, 1996). Finally, ex-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

posure to media coverage of an incident before an interview can also lead witnesses to recall incorrect details that were suggested or inferred (Crombag, Wagenaar, & van Koppen, 1996), and contact with other witnesses can similarly introduce systematic errors in memory (Gabbert, Memon, & Allan, 2003; Gabbert, Memon, Allan, & Wright, 2004; Eisen, Gabbert, Ying, & Williams, 2017; Loftus & Greene, 1980). We recommend that investigators *avoid* suggestive or leading interviewing practices and that they instruct witnesses *not to discuss* their accounts with or in front of one another. Whereas it is possible that discussions between witnesses could lead to an increase in reported details (e.g., Vredeveldt, Hildebrandt, & van Koppen, 2016), the dangers associated with contamination suggest that witnesses should be interviewed individually to preserve the independence of each statement. Investigators should also document whether a witness has spoken previously with other witnesses or has been exposed to media reports related to the incident.

Documenting Factors That Can Influence the Quality of Witness Accounts

It is also important that investigators note the conditions under which the witness may have viewed or interacted with the culprit, as certain factors can influence the likely quality of a witness's recollection (for a review see Meissner et al., 2007). Documenting such factors can aid both investigators and fact finders in assessing the likely reliability of a witness's memory. With respect to naturally occurring estimator variables, factors at the time of encoding such as low illumination (Wagenaar & van der Schrier, 1996), greater distance from the culprit (Loftus & Harley, 2005), and limited time of exposure can lead to poorer quality person descriptions (Sporer, 1992; van Koppen & Lochun, 1997; Yarmey, 1986; Yarmey, Jacob, & Porter, 2002). The presence of a weapon can draw attention away from the culprit's appearance (Fahsing et al., 2004; Fawcett, Russell, Peace, & Christie, 2011; Kocab & Sporer, 2016; Pickel, 1998, 1999). The consumption of alcohol or drugs by a witness can similarly reduce the amount of information provided (Flowe, Takarangi, Humphries, & Wright, 2016; Read, Yuille, & Tollestrup, 1992; Schreiber Compo et al., 2017; Yuille & Tollestrup, 1990). Extensive delays between encoding and the time of interviewing can diminish the amount of detail provided by a witness (Ellis, Shepherd, & Davies, 1980; Meissner, 2002; Tuckey & Brewer, 2003; van Koppen & Lochun, 1997). In contrast with the previous factors, prior familiarity with the culprit (i.e., an individual known to the witness) generally increases the accuracy of a witness's description and identification (Vallano, Steele, Slapinski, Briggs, & Pozzulo, 2019). Given the influence of these estimator variables on both the quality of person descriptions and subsequent attempts to identify the culprit from a lineup, it is recommended that investigators clearly document the presence of such factors in their report.

Evidence-Based Approaches for Interviewing Witnesses and Victims

Acquiring a complete, yet accurate, statement from the witness is critical to furthering an investigation. Considerable research has documented the most effective methods for interviewing a witness or victim (Dando et al., 2015; Fisher et al., 2014). In general, it is common for investigators to invite an *open-ended response* from the witness, followed by *specific probes* associated with key details such as the culprit's physical characteristics (e.g., height, build, age, race, sex, etc.), clothing, or any distinguishing characteristics (Brown et al., 2008; Launay & Py, 2015; Wise, Safer, & Maro, 2011). The use of open-ended, nonsuggestive questioning tactics (Clarke, Milne, & Bull, 2011; Walsh & Bull, 2010) is recommended for eliciting a complete narrative from the witness. Although the use of specific probes can increase the number of details provided, such details may come at the expense of lower accuracy of responding (Sauerland, Krix, van Kan, Glunz, & Sak, 2014). As such, caution should be used in moving to closed-ended or two-alternative questions and the use of suggestive/leading prompts should be avoided altogether.

Evidence-based interviewing protocols have been developed that both avoid the pitfalls of leading and suggestive questioning and enhance witness reporting by facilitating the retrieval of information from memory. One of the most notable and empirically validated protocols is the Cognitive Interview (Fisher & Geiselman, 1992). A robust literature has demonstrated the effectiveness of the Cognitive Interview for eliciting both detailed event narratives and person descriptions from cooperative witnesses (Memon, Meissner, & Fraser, 2010). Several instructional and mnemonic aspects of the Cognitive Interview appear to be particularly useful, including: (a) encouraging witnesses to "report all" of the information they can recall but not to guess about anything they are unsure of (e.g., Clifford & George, 1996); and (b) using context reinstatement by asking witnesses to close their eyes and think back to the event context (e.g., Smith-Spark, Bartimus, & Wilcock, 2017; Vredeveldt, Baddeley, & Hitch, 2012, 2014). Lab-based research indicates that the Cognitive Interview (compared with a standard police interview) increases the number of descriptors of the culprit and increases the chances that using the description can lead to finding the culprit's photo among a larger set of photos (Satin & Fisher, 2019). The increase in descriptors for the Cognitive Interview is typically quite large for correct details; and although a small increase in incorrect details has been noted across studies, the accuracy rate for the Cognitive Interview does not differ from that of a standard interview (see Memon et al., 2010). We encourage investigators to seek training in and adopt the Cognitive Interview protocol when interviewing witnesses and victims.

Another specific protocol for eliciting person descriptions, termed the Person Description Interview, significantly increases the quantity of person descriptors provided by witnesses (Demarchi & Py, 2009; Demarchi, Py, Groud-Than, Parain, & Brunel, 2013). The Person Description Interview incorporates two key instructions to the witness with respect to describing a person of interest: (a) to provide general information about the person before moving to specific featural aspects of the face, and (b) when describing the face to begin with the lower regions of the face (chin and lips) and to move up to the top regions (eyes and hair). Consistent with the Person Description Interview instructions, encouraging witnesses to provide more general, coarse-grained information during an interview can enhance the quantity of information absent a cost to accuracy (Brewer, Vagadia, Hope, & Gabbert, 2018).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

Finally, although witness descriptions are frequently collected via an oral interview conducted by an investigator, at times a witness may be asked to directly provide a written statement. There is mixed evidence with respect to how the format for eliciting a witness's recall might influence the quantity and quality of information provided, with some studies suggesting that oral interviews produce more information from witnesses than does written statements (Kraus, Zeier, Wagner, Paelecke, & Hewig, 2017; Sauerland & Sporer, 2011) and others finding no difference as a function of modality (McPhee, Paterson, & Kemp, 2014; Sauerland et al., 2014). A recently developed protocol, referred to as the Self-Administered Interview (see Hope, Gabbert, & Fisher, 2011), allows cooperative witnesses to self-generate high-quality descriptions of their experience. The Self-Administered Interview prompts witnesses to recall details of the event, including a person description of the culprit (e.g., hair, complexion, build, distinguishing features). To facilitate recall, the Self-Administered Interview incorporates the two key elements of the Cognitive Interview as previously described—a "report everything" instruction and a context reinstatement prompt. The Self-Administered Interview has been shown to significantly increase the quantity of person descriptors when compared with a standard free recall prompt, at a level comparable with that of the Cognitive Interview (Gabbert, Hope, & Fisher, 2009; Hope, Gabbert, Fisher, & Jamieson, 2014). Such an interview protocol is particularly useful when an incident involves many possible witnesses or victims and when such conditions could lead to significant delays in eliciting a statement from witnesses or victims.

Finally, we recommend that all interviews with witnesses should be video-recorded. Such an objective record of the interview will allow both investigators and fact finders the opportunity to review the information provided the witness and evaluate its evidential value. Importantly, studies suggest that investigators fail to accurately record or recall key details of statements provided in interviews (Kassin, Kukucka, Lawson, & DeCarlo, 2017; Lamb, Orbach, Sternberg, Hershkowitz, & Horowitz, 2000); thus, recording the interview with a witness provides an objective record of the information elicited, absent omissions or errors that may be introduced via the investigators' recollection of the interview. Lastly, the interview should be video-recorded from a perspective that captures both the investigator and the witness, as studies suggest that this perspective can enhance fact finders' evaluations of the evidence (Lassiter, 2010; Lassiter, Ware, Ratcliff, & Irvin, 2009; Ratcliff, Lassiter, Schmidt, & Snyder, 2006).

Instructions to Witnesses Following an Interview

Interviewers should instruct witnesses to not discuss the witnessed event or what they have told investigators with other potential witnesses in the case. As noted earlier, when eyewitnesses talk with each other about their memories, they can influence one another such that their subsequent individual memory reports can become contaminated with what others have recalled (for reviews, see Gabbert & Hope, 2013; Wright, Memon, Skagerberg, & Gabbert, 2009). Encountering a piece of misinformation from a cowitness about a facial feature can lead witnesses to later misidentify someone from a lineup who has that feature (e.g., Eisen et al., 2017; Zajac & Henderson, 2009). Moreover, hearing that a cowitness had made an identification from a lineup can increase the chances that a witness would also make an identification; hearing that a cowitness identification was made more confidently, as opposed to less confidently, can also increase the confidence that a witness expresses in the accuracy of their own identification (Levett, 2013). It is recommended (a) that witnesses should be cautioned to avoid discussing the case with others and (b) that investigators should refrain from sharing any information that other witnesses had previously provided. As with the lineup identification procedure (see Recommendation 7), we recommend that the entire prelineup interview be video-recorded.

Finally, interviewers should instruct witnesses not to conduct their own investigation of the crime. Increasingly, we are seeing that the first identification that witnesses make is the result of a self-directed search on the Internet, including social media sites. These searches by witnesses lack many of the protections of a well-conducted lineup; all faces viewed are possible suspects, there are no instructions reminding the witnesses that the culprit may not be among the faces they viewed, some faces may stand out more than others, witnesses might engage in these searches alongside cowitnesses, and there is no recording of confidence immediately after the identification. Once this initial identification of a suspect is made through an Internet search, it is not possible to conduct an uncontaminated identification procedure using better methods (see Recommendation 8).

Recommendation 2: Evidence-Based Suspicion

There should be evidence-based grounds to suspect that an individual is guilty of the specific crime being investigated before including that individual in an identification procedure and that evidence should be documented in writing prior to the lineup.

Conducting lineups in the absence of evidence-based reasons for suspicion is a risk factor for mistaken identification. In the parlance of eyewitness science, making an individual the focus of a lineup in the absence of evidence that the individual is likely to be the culprit (e.g., having only a hunch) contributes to a low base rate for culprit-present lineups (i.e., a high base rate for culprit-absent lineups). In the case of lineups, base rate refers to the rate for which the suspect in the lineup is guilty versus innocent. A proper lineup contains only one suspect, who might or might not be the culprit (see Recommendation 4). It follows from this structure of lineups that a mistaken identification of an innocent suspect cannot happen with a culprit-present lineup and, of course, an identification of the culprit cannot happen with a culprit-absent lineup (Wells & Turtle, 1986). Therefore, low base rates for culprit-present lineups (high base rates for culprit-absent lineups) create fertile ground for mistaken identifications of innocent suspects and reduce the chances of identifying the culprit. Moreover, culprit-absent lineups inflate the rate at which eyewitnesses identify known-innocent fillers (Smith, Wilford, Quigley-McBride, & Wells, 2019), thereby tainting that witness's credibility for any later lineup that might include the culprit.

The evidence-based suspicion recommendation derives from the observation that there are no laws or other mechanisms in place to prevent jurisdictions from making investigative decisions that result in extremely low base rates for culprit-present lineups (i.e., a high rate of culprit-absent lineups; Wells, 2006). In fact, the only study of actual lineups to estimate the base rate for culprit-present lineups in any jurisdiction (in this case the Houston, Texas Police

Department) yielded an estimate of a mere 35% (Wixted, Mickes, Dunn, Clark, & Wells, 2016). If this estimate is correct for Houston, then the suspect was innocent in 65% of their lineups. Another field study found that 40% of the lineups (in Northern California jurisdictions) had no prelineup evidence at all indicating that the suspect was the culprit; and for an additional 30% of the lineups there was minimal evidence (Behrman & Richards, 2005). Moreover, a national survey of U.S. law enforcement agencies reported that more than one third of the agencies stated that they needed no evidence at all or needed only a mere hunch that a person might be the culprit before placing that person in a lineup (Wise et al., 2011). Of course, the base rate is likely to vary from one jurisdiction to the next depending on the practices and policies in place (Wells, 1993). Nevertheless, at the time of this writing we know of no jurisdiction in the U.S. whose policies or written procedures require, urge, or even mention that there should be some form of concrete evidence against a person before conducting an identification procedure focused on that person.

It is unclear why so many crime investigators do not seem to be concerned about the problem with having little or no evidence before placing someone in a lineup. Perhaps this lack of concern stems from an assumption that an eyewitness would not pick an innocent individual and would only pick someone if they remembered the person committing the crime.

The Importance of Base Rates

There are many studies in the basic judgment and decision-making literature showing that people struggle to grasp the strong impact that prior probabilities and base rates have on test outcomes (e.g., Kahneman & Tversky, 1973). To illustrate, it is perhaps instructive to draw a close analogy between eyewitness identification testing and medical diagnostic testing. In medical diagnostic testing, it is common for medical organizations to issue guidelines about when to perform diagnostic tests versus forgo such tests. Consider, for example, the prostate-specific antigen (PSA) test for prostate cancer (see Vollmer, 2006). Although the PSA test is just as accurate for men under 30 as it is for men over 50, almost every positive PSA test result on men under 30 is a false alarm whereas only a small fraction of positive PSA test results on men over 50 are false alarms. Because the base rate (or prior probability) that an under-30 male will have prostate cancer is nearly zero, almost every positive result is a false alarm.

This same principle applies to eyewitness identification procedures. In the case of lineups, the base rate is the rate at which the suspect in the lineup is guilty versus innocent. More formal treatments of the (Bayesian) mathematics behind this problem are available (see Wells, Yang, & Smalarz, 2015; Wixted & Wells, 2017), but a simple version of the problem is presented here. Assume that the chances that an innocent suspect will be identified from a culprit-absent lineup is 6% and the chances that a guilty suspect will be identified from a culprit-present lineup is 60%. Assume as well that the long-term base rate for culprit-present lineups is 50% (and the culprit-absent lineup base rate is therefore 50%). Suppose now that 1,000 lineups were conducted (500 culprit-present and 500 culprit-absent). We would expect 300 identifications of guilty suspects (60% of 500) and 30 identifications of innocent suspects (6% of 500). In this example, 330 suspects are identified and 9.1% of these suspects (30/330 = .091) are innocent.

Now suppose that, instead of a 50% base rate, the base rate were lowered to 30% (300 culprit-present lineups and 700 culprit-absent lineups). Now, the 1,000 lineups would be expected to yield 180 identifications of guilty suspects (60% of 300) and 42 identifications of innocent suspects (6% of 700). The result is that 222 suspects are identified (180 + 42) and 18.9% of these (42/222 = .189) are innocent. In this 30% base-rate example, the percentage of identified suspects who are innocent more than doubles compared to when the base rate is 50%. Clearly, things get better if the base rate for the suspect being guilty is increased to 70%. At a base rate of 70%, the 1,000 lineups would yield 420 identifications of guilty suspects (60% of 700) and only 18 identifications of innocent suspect (6% of 300). In this example, 438 suspects are identified (420 + 18) and only 4.1% are innocent.

Notice in the earlier examples that the eyewitnesses themselves are performing just as well when the base rate is 30% as they are when the base rate is 70% (just as the PSA test performs as well when used on men of age 30 as it does on men of age 60). The difference is that the 30% base rate allows for many more false alarms than does the 70% base rate. Every time a culprit-absent lineup is conducted, there exists some probabilistic jeopardy for an innocent suspect. Therefore, minimizing the chances of presenting witnesses with culprit-absent lineups is one way to reduce the problem of wrongful convictions.

Even when the witness does not identify the innocent suspect in a culprit-absent lineup, they often identify a known-innocent lineup filler (Clark & Wells, 2008; Wells & Lindsay, 1980; Wells & Olson, 2002; Wells et al., 2015). Most filler identifications are made with low confidence, clearly signaling their error-prone nature (Wixted & Wells, 2017); however, some are made with higher confidence. Suppose, for example, an eyewitness is shown a culprit-absent lineup and identifies a known-innocent filler. Later, police receive information about who the actual culprit is. It is too late to undo the fact that showing the eyewitness a culprit-absent lineup led the witness to identify a known-innocent filler as the culprit. This prior identification of a known-innocent filler makes the prosecution of any newly identified person in a later lineup substantially more difficult for prosecutors. In this sense, culprit-absent lineups not only create risk for innocent suspects but also elevate rates of filler identifications that, in turn, undermine eyewitnesses' credibility on any later identification opportunities that could involve a culprit-present lineup (Wells, Steblay, & Dysart, 2012).

What Is Evidence-Based Suspicion?

By evidence-based suspicion, we mean that there is articulable evidence that leads to a reasonable inference that a particular person, to the exclusion of most other people, likely committed the crime in question. As with other standards used in the legal system (such as reasonable suspicion or probable cause), there is no precise probability associated with the concept of evidence-based suspicion. However, a mere hunch is not evidence-based suspicion. Moreover, merely fitting a general description that the witness gave of the culprit (e.g., young male, mid-20s, dark hair, normal build) is not evidence-based suspicion as it could be applied to large numbers of people. Nor can this notion of

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

evidence-based suspicion be based on backward reasoning in which a pick of that person in the lineup is used retroactively to justify the placement of the person in the lineup. In other words, evidence-based suspicion is something that must be established prior to the lineup.

Not all articulable evidence qualifies as good evidence that connects a specific suspect to a specific crime. However, there are many possible examples of what could qualify as evidence-based suspicion for purposes of justifying the placement of an individual in a lineup identification procedure. Examples include:

- A unique fit to a specific description that was given by the eyewitness (e.g., blue teardrop tattoo under left eye; moon shaped scar on chin);
- Self-incriminating statements;
- Being in possession of materials linked to the crime along with a fit to the general physical description given by the eyewitness;
- Known to be in the area of the crime around the time of the crime along with a fit to the general physical description given by the eyewitness;
- Physical evidence at the crime scene linked to the person along with a fit to the general physical description given by the eyewitness;
- A unique pattern to the crime that is known to be associated with a particular offender along with a fit to the general physical description given by the eyewitness.

These examples of evidence-based suspicion are certainly not exhaustive. But they capture the idea that the evidence should be articulable, not based on mere hunch, and lead to a reasonable inference that there is individuating evidence that makes this person, to the exclusion of most other people, a reasonable candidate to be the one who committed the crime in question.

In contrast, articulable evidence that fails to link the suspect to the specific crime for which an identification is sought does not qualify as evidence-based suspicion. Some examples include:

- A search of police records reveals that the suspect was convicted of a similar crime in the same jurisdiction, has been released from prison, and is now living in the neighborhood where the crime was committed.
- There are multiple witnesses to a crime and the identification of the suspect made by the first witness is used to establish evidence-based suspicion for the remaining witnesses. If there was no articulable evidence for the first witness then the first lineup should never have been conducted.
- The suspect resembles a composite sketch or rendering of the culprit made with the assistance of the witness. This type of evidence does not clear the threshold of reasonable, articulable suspicion linking the suspect to the crime under investigation because composites do not reliably represent a recognizable representation of the culprit (Kovera, Penrod, Pappas, & Thill, 1997).
- A suspect who was apprehended in the vicinity of one crime happens to match the description of the culprit not only for that crime but also for several other similar crimes recently committed elsewhere in the community. If a witness to the crime committed in the vicinity of the suspect's apprehension does not identify the suspect, then there is no evidence to support placing that suspect in lineups shown to witnesses to the other similar crimes committed elsewhere in the community. The nonidentification fails to establish any link between the suspect and the other similar crimes.

Again, this list of articulable evidence that does not meet criteria for evidence-based suspicion is not exhaustive. However, these examples illustrate that the evidence supporting the placement of a suspect in an identification procedure must be evaluated for whether it actually provides a nexus between the suspect and the crime witnessed.

Final Comments on Evidence-Based Suspicion

The medical field's understanding of the impact of base rates on medical diagnostic test outcomes is far ahead of the legal system's understanding of the impact of base rates in eyewitness lineup test outcomes (Wells et al., 2015). Furthermore, the concern about base rates in eyewitness identification might be even more important than are base rate concerns in medical testing because diagnostic medical tests can be repeated to confirm reliability of the result, or a different type of test can be performed to look for convergence of results. An eyewitness identification test, in contrast, cannot be repeated with that same witness and same suspect without being contaminated (see Recommendation 8); it is important to ensure that the chances of presenting eyewitnesses with a culprit-absent lineup are not unduly high.

Although our discussion of the importance of having evidence-based suspicion has been centered on the chances that an innocent suspect will be identified, there is an additional reason to be concerned about presenting eyewitnesses with a culprit-absent lineup. Even if the eyewitness does not identify an innocent suspect, culprit-absent lineups strongly increase the chances that the eyewitness will identify a filler (Wells, 1984; Wells et al., 2015). When an eyewitness identifies a filler, it "burns" the credibility of that eyewitness for purposes of any later identification (e.g., Wells et al., 2012). Suppose, for example, an innocent suspect was placed in a lineup, the eyewitness picks a filler, and investigators later discover evidence-based suspicion against a new suspect. Can they simply bring the eyewitness back and show the eyewitness a new lineup with the new suspect? The empirical data indicate that eyewitnesses who identify a filler from a culprit-absent lineup are highly error prone on any later lineup, even if that later lineup includes the culprit (Smalarz, Kornell, Vaughn, & Palmer, 2019). Moreover, research indicates that giving disconfirming feedback to witnesses who identify a filler reduces performance on subsequent identification tests (Palmer, Brewer, & Weber, 2010).

An evidence-based suspicion standard could be implemented easily by requiring detectives to present their proposal for conducting a lineup to a supervisor of detectives. The supervisor of detectives could then question the detective about why this lineup is being conducted with this particular person as its focus (Wells, 2006). The detective should be able to point to some concrete evidence that could lead to a reasonable inference that this person should be suspected of being the culprit in question; if not, a supervisor of detectives could suggest instead that the detective investigate further so as to have more confidence that the subject of the lineup is the culprit.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

Although a lineup should be conducted only after establishment of evidence-based suspicion, eyewitness memory can fade with the passage of time. Hence, a lineup should be conducted as soon as possible after establishing evidence-based suspicion.

Recommendation 3: Double-Blind (or Equivalent)

Lineups should be conducted using a double-blind procedure (i.e., neither the administrator nor the witness should know who the suspect is in the lineup) or an equally effective method of preventing the lineup administrator from inadvertently influencing the witness.

A lineup administration is a social interaction between a witness and an administrator. Like in other social situations, interpersonal expectancies operate in the context of a lineup administration. Why is the social interaction aspect of lineup administration concerning? When someone has an expectation about how another is likely or ought to behave, this expectation can cause the person with the expectation to behave differently toward the target of the expectation. This change in the expectation-holder's behavior elicits the very behavior that was expected from the target (Harris & Rosenthal, 1985; Rosenthal, 2002; Snyder & Swann, 1978). The social interaction that takes place during the administration of a lineup is not immune from this interpersonal expectancy phenomenon. There is no presumption that the influence of the lineup administrator is intentional or even that the lineup administrator or witness is aware of the influence.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

Lineups-As-Experiments

A lineup is a test of the hypothesis that the person whom the police suspect is in fact the culprit of the crime. The lineup administrator is fundamentally an experimenter who is conducting a procedure to test this hypothesis (Wells & Luus, 1990). Because people tend to test hypotheses in a way that will confirm their expectations (e.g., Klayman & Ha, 1987; Skov & Sherman, 1986), a lineup administrator, like any other experimenter, should follow protocols that will prevent them and their expectations from influencing the results of their tests. Double-blind testing, in which the lineup administrator does not know which person is the suspect and which are merely fillers (i.e., a blind administrator), is the best way of ensuring that any information that administrators have about which lineup member is the suspect will not influence the witnesses' behavior, including any identification decision they might make or their confidence in that decision (see also Recommendation 6 for a discussion of how double-blind administration eliminates the opportunity for postidentification feedback that could influence witness confidence). Double-blind testing can also prevent administrators' expectations from influencing their reports of witnesses' behaviors during the procedure. In contrast, single-blind lineup administration, in which the administrator knows which lineup member is the suspect and which are fillers (i.e., nonblind administrator), allows for the possibility that the administrator will communicate the identity of the suspect to the witness through intentional or unintentional verbal or nonverbal behaviors. In a single-blind lineup procedure, the eyewitness does not know which person is suspected of being the culprit and which ones are fillers, but the lineup administrator knows. In a double-blind lineup procedure, neither the eyewitness nor the lineup administrator knows which person is suspected of being the culprit and which are fillers.

The double-blind recommendation is primarily focused on keeping knowledge about the suspect from the administrator of a lineup so that this knowledge cannot influence the administrator's behavior while conducting the identification procedure. However, the purpose of the recommendation is to keep anyone who knows which lineup member is the suspect from influencing the witness. Thus, there should be no officers (e.g., the lead detective) in the room where the identification procedure is conducted who know which lineup member is the suspect, even if they are not the officer administering the procedure. Moreover, if there are multiple witnesses, a different blind administrator should conduct the lineup with each witness because conducting a procedure with one witness may provide an administrator with clues about which lineup member is the suspect, which might then influence how that administrator interacts with the next witness while administering the identification procedure (Douglass, Smith, & Fraser-Thill, 2005).

It is equally important to keep information about which person is the suspect and which ones are fillers from the witness. It might seem odd to explicitly warn against letting a witness know who the suspect is before they make an identification decision. In practice, however, we find it not uncommon for circumstances surrounding the identification procedure to alert the witness to which lineup member is the suspect. For example, the witness may be tipped off to who the suspect is after being presented with multiple photo arrays that share only one lineup member (the suspect) in common, which is one of the many reasons for our recommendation to avoid repeated lineup procedures (Recommendation 8).

The recommendation for double-blind administration of lineups was included among the original four recommendations made in the previous scientific review paper (Wells et al., 1998). At the time that article was written, however, there were no studies that directly tested whether a lineup administrator's knowledge of which lineup member was the suspect influenced witness identifications. Without studies directly testing the effects of double-blind administration of lineups, the recommendation was made based on generalizations from basic studies on experimenter expectancy effects (Harris & Rosenthal, 1985) and early research suggesting that positive feedback to witnesses after they choose the suspect increases their confidence in that choice (Wells & Bradfield, 1998). Since the previous scientific review paper was written, researchers have conducted a number of studies demonstrating that the single-blind administration of lineups increases the likelihood that witnesses will identify the suspect (for a review, see Kovera & Evelo, 2017), irrespective of whether the suspect is the culprit (Charman & Quiroz, 2016; Greathouse & Kovera, 2009) or an innocent suspect (Charman & Quiroz, 2016; Greathouse & Kovera, 2009; Zimmerman, Chorn, Rhead, Evelo, & Kovera, 2017).

Paradigms for Examining Lineup Administrator Influence

Scholars have developed several paradigms to examine the effects of administrator influence on witness decisions. In one paradigm, which has been termed the *steering paradigm* (Kovera & Evelo, 2017), the lineup administrator is a confederate of the experimenter who intentionally engages in behaviors that steer that witness toward the suspect (e.g., Rhead, Rodriguez, Korobeynikov, Yip, & Kovera, 2015) or encourages the witness to make an identification (Clark, Brower, Rosenthal, Hicks, & Moreland, 2013). However, the influence of administrators on witnesses need not be intentional. In the *cue-disruption paradigm*, all administrators know who the suspect is but half of them are prevented from sending cues (whether intentionally or unintentionally) to the witness during the administration of the lineup. In one study, for example, the contact between the administrator and the witness was limited by having the administrators stand behind the witnesses while they viewed a photo-array. Witnesses were less likely to identify the suspect when the administrator stood behind them than when the administrator sat in front of or beside the witness (Haw & Fisher, 2004). In the *double-blind paradigm*, participants are randomly assigned to be either witnesses or lineup administrators; half of the lineup administrators are told who the suspect is and the other half are not (Phillips, McAuliff, Kovera, & Cutler, 1999). These participant administrators then present the photo-array to the participant witnesses. Across all paradigms, when administrators know who the suspect is and are not prevented from sending cues to the witness, witnesses are more likely to choose the suspect from the lineup, whether the suspect is the culprit or not (Kovera & Evelo, 2017).

These studies tell us that changes in the behavior of administrators during the administration of the lineup are responsible for this increase in witness picks of the suspect. In the steering paradigm studies, the administrators' behaviors were intentionally manipulated to steer the witness toward the suspect and away from fillers (e.g., Rhead et al., 2015). In double-blind paradigm studies, observers reported that nonblind administrators placed more pressure on witnesses to choose someone from the lineup than did blind administrators (Greathouse & Kovera, 2009) and that pressure was directed toward choosing the suspect rather than a filler (Zimmerman et al., 2017). Nonblind administrators were more likely to directly ask witnesses about the suspect than were blind administrators (Zimmerman et al., 2017). Nonblind administrator influence can be nonverbal as well; nonblind administrators are also more likely than blind administrators to smile when a witness is looking at the suspect rather than a filler (Charman & Quiroz, 2016; Zimmerman et al., 2017).

These differences in behaviors between blind and nonblind administrators affect which photo witnesses choose from lineups, not whether they make a choice at all. Witnesses are equally likely to choose someone from a lineup, irrespective of whether the lineup administrator knows who the suspect is (Greathouse & Kovera, 2009; Kovera & Evelo, 2017). The increase in witness identifications of the suspect from single-blind lineup administrations appears to be the result of witnesses who would have identified a filler (and do so under blind administration) identifying the suspect instead due to influence from the nonblind administrator (Kovera & Evelo, 2017). This pattern of findings, replicated in a number of studies (Charman & Quiroz, 2016; Greathouse & Kovera, 2009; Kovera & Evelo, 2017), is known as the filler-to-suspect shift and provides compelling evidence that single-blind lineup administration allows administrators to transmit information about who the suspect is to witnesses, even if unintentionally.

Double-Blind Administration Helps Prevent Postidentification Feedback

In addition to affecting witnesses' identification decisions, single-blind lineup administration allows administrators to provide feedback to witnesses about their decisions. Nonblind administrators react to witness identifications in ways that send information to witnesses about whether their choice was "correct" (i.e., an identification of the suspect; Charman & Quiroz, 2016; Garrioch & Brimacombe, 2001). Two decades of research supports the conclusion that providing feedback to witnesses that they identified the suspect increases their confidence in the accuracy of their decision, especially among eyewitnesses who have made a mistaken identification (Steblay, Wells, & Douglass, 2014; see Recommendation 6 for a more complete discussion of this research). This confirming feedback effect attenuates the relationship between confidence and accuracy (Bradfield, Wells, & Olson, 2002), rendering witnesses' reports of their confidence useless for judging their accuracy (Wixted & Wells, 2017). In addition to preventing administrators from providing feedback that will influence witnesses' reports of their confidence, double-blind administration will also prevent other unwelcome effects of feedback such as the contamination of witnesses' memory for the conditions under which they witnessed the crime (Steblay et al., 2014), the impairment of witness memory for the culprit (Smalarz & Wells, 2014a), and lessening the ability of jurors to differentiate between accurate and inaccurate witnesses (Smalarz & Wells, 2014b).

Double-Blind Administration Helps Ensure Full and Accurate Reports

Knowing who the suspect is may also influence what information administrators record about witnesses' behavior during the identification procedure. Even though we recommend video-recording the lineup administration (see Recommendation 7), a 2013 survey indicated that most lineups in the U.S. (>75%) are not video-recorded (Police Executive Research Forum, 2013). In cases that fail to video-record, the only contemporaneous record of what happened during the procedure is information memorialized by the administrator. Lineup administrators often fail to make a record of the verbatim statement of the eyewitness but instead will make a note of the gist of what the eyewitness said. It is also possible that knowing who the suspect is may change how administrators assess and record witnesses' choices from lineups. If so, when witnesses make tentative identifications (e.g., "I don't know. I think it may be Number 4, but I'm not certain."), administrators who know that the witness is talking about a suspect may record a positive identification of the suspect whereas administrators who know that the witness is talking about a filler may record the very same behavior as a nonidentification or rejection (Rodriguez & Berry, 2014). Moreover, lineup administrators' interpretations of ambiguous eyewitness statements and administrators' perceptions of the witness are biased by whether the lineup administrator is blind or not blind (Charman, Matuku, & Mook, 2019).

Although there are limited empirical data that directly bear on the effects of administrators' knowledge of who the suspect is on their reports, evidence continues to mount that forensic examiners' expectancies influence their evaluations (for a review, see Kassin, Dror, & Kukucka, 2013). In addition, there are data from both

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

laboratory and field studies suggesting that administrators who are nonblind record witness choices differently than do administrators who are blind. Indeed, for 5 years in Queens County, NY, the District Attorneys' Office recorded choices made by witnesses from live single-blind lineups. Supposedly, choices were only recorded when the administrators judged that it had been made with a high degree of confidence and was not tentative (as reported in Mecklenburg, 2006). But in these Queen's county cases, it was nonblind administrators who made decisions as to whether to report an affirmative response as an identification or whether to dismiss it as something else. In other words, the administrators in the Queens County cases knew that these were filler picks. This procedure resulted in a very low rate of *reported* filler identifications (between 0.56% and 5.62%). Field studies for which the lineups were conducted double-blind, however, report filler identification rates that are much higher 11%–15% (see Klobuchar, Steblay, & Caligiuri, 2006; Wells, Steblay, & Dysart, 2015). This difference in filler identification rates could represent a difference in reporting that derives from whether the lineup was double-blind.

In other field data from the Evanston Police Department in Illinois, reports from double-blind lineups were more likely to involve verbatim reports of witness statements than were reports from single-blind lineups (83% vs. 39%, Steblay, 2011). The interpretation of the Evanston data is problematic because only double-blind administrators were instructed to record what words the witnesses used to make their identification whereas single-blind administrators were not.

Controlled experiments have tested how administrators make records of the behavior of eyewitnesses as a function of whether the administrator of a lineup was blind and whether a confederate-witness chose the suspect or a filler (Rodriguez & Berry, 2014, 2019). Although double-blind administrators were just as likely to report that witnesses had made a positive identification when the witness identified a filler as when the witness identified the suspect, single-blind administrators were more likely to report incorrectly that witnesses who identified a filler had not made an identification. Moreover, when nonblind administrators recorded the confidence reported by witnesses, independent coders who were blind to condition judged the confidence levels of those who identified suspects to be higher than those who identified fillers even though the confederate-witness expressed the same level of confidence in both types of identifications (Rodriguez & Berry, 2019).

These effects of single-blind lineup administration on witnesses' identification decisions, their confidence, and administrators' reporting behavior support the use of double-blind procedures when collecting eyewitness identification evidence. The U.S. Department of Justice (2017) and the National Research Council (2014) have made similar recommendations. Nevertheless, many jurisdictions have yet to put this procedure, or a similar alternative, into practice (Kovera & Evelo, 2017).

Practical Issues in Implementing Double-Blind Lineups

Some resistance to double-blind lineup administration has come from a limited-resources argument that certain police departments are so small that every officer in the department knows the identity of the suspect. At the state level, this concern has not been found in actual practice. In 2002, New Jersey became the first state to mandate double-blind administration, and state officials have reported no problems implementing this policy. In cases of very small police departments, for example, cooperative agreements were created to loan officers to nearby departments for the purpose of conducting double-blind identification procedures. Other states, from Florida to California, have similarly reported no problems with conducting double-blind lineups.

Hence, we recommend double-blind lineup administration and believe that actual practice has proven it to be viable for all jurisdictions. Nevertheless, in theory any procedure that prevents the possibility of a nonblind lineup administrator influencing the eyewitness could be used. With photo lineups for example, it is possible to use a laptop computer with software that delivers prelineup instructions, randomizes and presents the photo lineup, records any identification decision from mouse clicks, and collects a confidence statement from the eyewitness. With such software, the eyewitness can self-administer the photo lineup without anyone else present in the room, thereby guaranteeing that there could not have been lineup administrator influence over the eyewitnesses' identification or confidence statements. We recommend that video-recording be used with the laptop procedure just as it is with a double-blind administrator (see Recommendation 7).

A low-tech alternative to the self-administered laptop procedure for photo lineups is the self-administered envelope method. With the envelope method, a photo lineup is prepared with clearly numbered photos and the page should clearly state the options (to identify one of the photos, indicate "not there," or indicate do not know in a way that is parallel to the instructions in Recommendation 5). The page should also include a confidence question (see Recommendation 6). A photo lineup with these items should be placed in a large envelope and sealed. After giving complete instructions to the eyewitness (see Recommendation 5), the lineup administrator should tell the eyewitness that the photos are inside of the envelope. Of course, when using the self-administered envelope method as an alternative to the double-blind method, the instruction that the lineup administrator does not know which lineup member is the suspect cannot be used. The witness should be instructed to make an identification decision by either circling the photo of the person they believe to be the culprit or circling the "none" or "do not know" option below the photos. The witness should be instructed to place the photo-lineup and responses back in the envelope before opening the door to tell the lineup administrator that she or he has finished. The witness should not be handed the envelope until the lineup administrator is prepared to leave the room. The outside of the envelope should again tell the eyewitness to open the envelope and to view the photos only after the officer has left the room and to replace the photos and the answers to questions back in the envelope before opening the door to let the officer know that they are finished. Only when the witness has confirmed that the identification decision and confidence statement have been completed and placed back in the envelope should the officer reenter the room and examine the results.

This envelope method could be adapted for sequential presentation of the photos, with photos placed individually in smaller, numbered envelopes and instructions to look at each photo in numerical order and record an identification decision and a con-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

fidence judgment before replacing the photo in its envelope and proceeding to the next envelope. Backloading of the lineup could be achieved by placing additional envelopes with blank photo pages in the later numbered envelopes, with an instruction in the first envelope used for backloading that the lineup procedure is complete and witnesses should return all materials to the large envelope and let the administrator know that they are done. As with all identification procedures, the self-administered envelope procedure should be video-recorded to ensure that the witness follows the instructions given.

In addition to concerns about limited resources making it difficult to implement double-blind procedures, some have objected to their adoption because of the loss of correct identifications associated with double-blind administration (Clark, 2012a, 2012b). Although it is true that double-blind procedures can reduce both correct and mistaken identifications, they do so by eliminating the opportunity for administrators to cue the witness to which lineup member is the suspect. Given that the legal system requires that an eyewitness identification be based on the independent memory of the witness (*Perry v. New Hampshire*, 2012), the loss of an identification obtained through administrator influence should not be a concern. Indeed, some scholars have termed these correct identifications obtained through suggestive procedures to be "illegitimate hits" (i.e., correct identifications that are not based in the witness's memory but instead a product of the cues received from an administrator; Wells et al., 2012). Thus, double-blind procedures serve to protect suspects' rights to due process.

Recommendation 4: Lineup Fillers

There should be only one suspect per lineup, and the lineup should contain at least five appropriate fillers who do not make the suspect stand out in the lineup based upon physical appearances or other contextual factors such as clothing or background.

Recommendation 4 concerns what might be considered the most widely known problem that can afflict lineups, namely a lineup that is constructed in a way that makes it obvious which member is the suspect. This idea is the source of scores of cartoons and jokes about the perceived failings of criminal justice, such as one that depicts a person embedded in a lineup composed of a dog, cat, refrigerator, and a microwave oven. Nevertheless, jocular treatments of biased lineups hide a serious problem and this problem has much greater complexity than meets the eye.

The problem of biased lineups is one of the oldest in the scientific study of eyewitness identification. In fact, the first published experiment on eyewitness identification that manipulated the presence versus absence of the culprit in the lineup (a now routine feature of eyewitness identification experiments) was an experiment in which the researchers also manipulated the lineup fillers to be either similar or dissimilar to the suspect (Lindsay & Wells, 1980). Not surprisingly, in culprit-absent lineup conditions the use of high-similarity lineup fillers strongly reduced mistaken identifications of the innocent suspect compared with the use of low-similarity fillers. In culprit-present conditions, the use of these same high-similarity fillers had only a minor impact on accurate identifications of the culprit relative to the use of low-similarity fillers. This pattern of results, showing that using low-similarity fillers increases the chances of mistaken identification of an innocent suspect, has been repeatedly replicated (Fitzgerald, Price, Oriet, & Charman, 2013).

Despite the relative ease of replicating the basic finding that low-similarity fillers increase the risk of mistaken identification of an innocent suspect, there is not total agreement among eyewitness scientists regarding the best strategy for choosing fillers to serve in a lineup. There are two primary strategies for selecting fillers (Luus & Wells, 1991). One strategy uses the verbal description of the culprit that the eyewitness provided (e.g., "White male, mid-20s in age, about 5 feet 10 inches tall, short dark hair, no facial hair, medium build"). This method of selecting fillers is called the *match-to-description strategy*. The alternative strategy, called the *resemble-suspect strategy*, involves selecting fillers who physically resemble the suspect. The resemble-suspect strategy can be problematic because it has no criterion or "stopping point" for determining how similar the fillers should be, at times resulting in lineup fillers who are too similar and leading to a different problem than a biased lineup (Luus & Wells, 1991). In effect, extremely high similarity creates a lineup of near-clones, thereby making it too difficult to identify the culprit from a culprit-present lineup. The match-to-description strategy, in contrast, has a natural stopping point (the description) and does not risk creating such high levels of similarity between fillers and the suspect that would interfere with obtaining accurate identifications of the culprit (Luus & Wells, 1991). An experiment comparing the two strategies to a biased (low similarity) lineup showed the two strategies to be equally effective in reducing innocent suspect identifications; however, the resemble-suspect strategy produced a reduction in accurate identifications of the culprit, whereas the match-to-description strategy did not (Wells, Rydell, & Seelau, 1993). These findings have been replicated (Juslin, Olsson, & Winman, 1996). Other studies have shown either no detrimental effect on culprit identifications from using the resemble-suspect strategy and no evidence that it made the innocent suspect stand out (Tunnicliff & Clark, 2000), or no advantage one way or the other for match-to-description versus suspect resemblance strategies (Darling, Valentine, & Memon, 2008). But a large-scale study comparing description-matched fillers to suspect-matched fillers showed clear evidence favoring description-matched fillers (Carlson et al., 2019).

In a meta-analysis of the data on filler similarity, lineups classified as high similarity produced a reduction in culprit identifications relative to low-similarity lineups but not relative to moderate similarity lineups (Fitzgerald et al., 2013). As noted in the meta-analysis, the categorization of low, medium, and high similarity lineups reflected relative rather than absolute levels of similarity (the concept of similarity could not be defined in absolute terms, with clear criteria for each level of sameness). In a more recent study, fillers who resembled the suspect were selected from either an extremely large database of faces (which produced very high similarity fillers) or a more modest sized database (which produced more moderate levels of similarity; Bergold & Heaton, 2018). Compared with the more modest size database of faces, using the large database of faces for selecting fillers resulted in a reduction in accurate identifications of the culprit by producing too much similarity between the fillers and the suspect. Overall, the data suggest that using the resemble-suspect strategy could produce too much similarity between the suspect and the fillers such that it interferes with identifications of the culprit, especially when

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

fillers are selected from large databases. This problematic condition could become more prevalent as the ability to rapidly search large databases of faces becomes increasingly more common.

A Blended Approach for Choosing Fillers

The net result of these complex problems is that the science has not yet been able to specify what the optimal level of similarity of fillers to the suspect ought to be and thus, at this time, there is no single strategy or formula for selecting fillers to be used in a lineup. Nevertheless, there are generally accepted principles regarding how fillers should be selected for a lineup, and they tend to involve a blend of the match-to-description and resemble-suspect strategies. First, there is general agreement among experts for a *minimal* requirement that fillers should fit the description that the eyewitness gave of the culprit. A failure to match the witness's description might introduce a serious bias even though lineup members may look very similar. For example, a witness might describe the culprit as "male, 40s, shaven head, striking blue eyes, very solid build around the neck and shoulders." A subsequent lineup could have an array of people who are very similar in appearance; however, if only one of the lineup members had blue eyes, that individual becomes distinct within the array in that his face will match the memory of the perpetrator more so than will the fillers.

An exception to this general match-to-description principle should be made when the suspect does not fit the witness's description. For example, a person might become a suspect for reasons other than his or her appearance. Furthermore, if the description of the culprit mentioned a moustache, but the suspect does not have a moustache, then the fillers also should not have a moustache. In other words, if there is a discrepancy on some physical feature between the eyewitness's description of the culprit and the appearance of the suspect, the fillers should match the suspect's appearance (rather than the witness's description of the culprit) on that feature.

Another situation in which match-to-description is inadequate is when the description is vague, general, or sparse. For example, the description "young White male" is inadequate. In this case, the fillers should match the suspect on basic "default" characteristics such as facial hair, hairstyle, and general body build (Lindsay, Martin, & Webber, 1994). Hence, the match-to-description method for selecting fillers should be used only if the description is complete.

Sometimes a suspect has a unique feature such as a tattoo or a scar, making it very difficult to find a filler who matches on that unique feature. There are two general approaches to dealing with this issue, made easier by advances in technology. One approach is to duplicate this feature on the fillers, which could be done electronically in the case of photo lineups. The other approach is to cover the unique feature on the suspect and then place that same cover on each of the fillers at the same location on their bodies. Both the "duplicate" and "cover" approaches appear to be equally effective (Colloff, Wade, & Strange, 2016). However, these alterations must be done in a way that does not make the suspect stand out in a mock witness test (see next section).

Another complication in selecting fillers can occur when a person becomes a suspect based on resemblance to a facial composite (e.g., forensic sketch or computer-generated face) or a surveillance image (although see discussion of the problems associated with using culprit-match to a composite in Recommendation 2: The Evidence-Based Suspicion Recommendation). If the person became a suspect based on resemblance to a composite or to a surveillance image whereas the fillers were chosen based merely on their match to the witness's verbal description, then there is a risk that the suspect will stand out. Hence, in such cases, fillers for a lineup need to be chosen based on their similarity to that same composite or surveillance image rather than chosen based on the verbal description given by the eyewitness (see Wixted & Wells, 2017).

"Mock Witness" Testing

There is broad agreement that the lineup that is created should be able to pass a "mock witness" test. A mock witness test is one in which a large number of people are individually given the description that the witness had given of the culprit, then shown the lineup and asked which person they think is the suspect. The ideal outcome from such a test would be if the suspect were selected by these mock witnesses only $1/N$th of the time, where N is the nominal number of lineup members. So, for a six-person lineup, a good outcome would be if the suspect were picked one sixth of the time. Consider again the witness description "male, 40s, shaven head, striking blue eyes, very solid build around neck and shoulders" and the suspect is the only one with blue eyes. This lineup is likely to result in most mock witnesses choosing the suspect rather than spreading their choices across the lineup members. We are *not* suggesting that police should be required to conduct a mock witness test on each lineup they create. Instead, we believe that a conscientious and objective detective would have a good sense of whether the lineup was fair without conducting a mock witness test with a large number of people. However, we do recommend that a nonblind officer building the lineup ask at least one or two other people (blind as to which person is the suspect) to review the witness description and evaluate the lineup with respect to whether it would pass a mock witness test. We also recommend that every lineup report include a written record of how the fillers were selected for the lineup.

Mock witness tests have been around since the 1970s (Doob & Kirshenbaum, 1973) and a number of different statistics have been developed to estimate lineup bias from mock witness results (e.g., Malpass, 1981; Tredoux, 1998; Wells, Leippe, & Ostrom, 1979). Mock witness measures tend to predict choices of eyewitnesses from culprit-absent lineups (Tredoux, Parker, & Nunez, 2007), though mock witness tests do have limits (see Wells & Bradfield, 1999). One of these limits is that a mock witness test is insensitive to whether the level of similarity between fillers and the suspect is too high, so high that it would likely harm rates of accurate identifications if the suspect is the actual culprit. For example, a lineup of clones would produce a good result from a mock witness test (one sixth of choices are of the suspect), even though it would not be a good lineup in that most witnesses would not be able to distinguish between the culprit and the fillers. In addition, because mock witness tests use the description of the culprit provided by the eyewitness, a mock witness test can appear perfectly fair if the description is sparse whereas that same lineup can appear quite unfair if the description is detailed (Mansour, Beaudry, Kalmet, Bertrand, & Lindsay, 2017).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

Background, Clothing, and Other Contextual Factors

The physical characteristics of the fillers are not the only factors that can make the suspect stand out in a lineup. In photo lineups, for example, the background of the photos, the size or brightness of the images, and the source of the photo could make the suspect's photo stand out from the others. Backgrounds are relatively easy to fix on photos with modern editing software. Similarly, clothing can easily be eliminated from headshots with editing software. Sometimes, a suspect's photo is from a different source (e.g., employment ID, social media, Department of Motor Vehicles driver's license) than a police department's usual source for photo lineups (e.g., mugshots), and this can make the suspect's photo stand out. This type of discrepancy will require either careful electronic editing or perhaps going to the same source (e.g., Department of Motor Vehicles drivers' licenses) to locate appropriate fillers.

It is not the case that every aspect of background, clothing, and other features must be exactly the same. The critical issue is whether the suspect stands-out in the context of the other fillers. So, for example, if the background of every photo was different from every other photo in the lineup, then perhaps the suspect's photo would not be any more distinct than other lineup members.

Articles of clothing (e.g., hats, shirts, etc.), however, could be of special concern. Specifically, if there are reasons to believe that the clothing worn by the suspect is similar to that worn by the culprit, then (a) every filler needs to be clothed that same way, (b) the suspect needs to have his or her clothes changed to blend in with the fillers, or (c) each lineup member's clothing must be obscured from the view of the witness. For cases in which the identification of clothing is thought to have potential probative value, a separate identification procedure involving *only* the papers of clothing (a clothing lineup) could be conducted independently of the identification of the suspect (see Lindsay et al., 1994; Lindsay, Wallbridge, & Drennan, 1987).

The Single-Suspect and Minimum of Five Fillers Requirement

A central feature of this recommendation is that the lineup should have only one suspect. There are several reasons why there should be only one suspected person in the lineup with the remainder having the status of being known-innocent fillers. Suppose, for example, that a lineup were composed entirely of suspects and no fillers. All-suspect lineups have been likened to a multiple-choice test in which there is no wrong answer (Wells & Turtle, 1986; Wixted & Wells, 2017). The value of having known-innocent fillers is that unreliable eyewitnesses are likely to err on a filler rather than on an innocent suspect (assuming that the lineup is composed of good fillers).

Consider again Table 1 of the current article in which we displayed the outcomes of lineups in actual cases. These were all single suspect lineups in which the suspect was embedded among fillers. Notice that these eyewitnesses to serious crimes identified fillers approximately 37% of the time they made an identification. If everyone in those lineups had been a suspect, all 37% of these would be mistaken identifications of innocent suspects who would then be subject to arrest and possible prosecution. These field data reinforce the dangers of having multiple suspects in a lineup, an issue that was first documented over 30 years ago through statistical proofs using data from eyewitness identification experiments (Wells & Turtle, 1986).

The recommendation that there should be at least five known-innocent fillers (thereby creating a six-person lineup) for a single suspect is somewhat arbitrary. At a theoretical level, we can say that an innocent suspect is better protected from mistaken identification with a six-person lineup than a five-person lineup, which is better protection than a four-person lineup, and so on, as long as other things are equal (e.g., how good the fillers are). However, there are diminishing returns (in terms of restricting mistaken identifications of innocent suspects). After all, in terms of protecting an innocent suspect, an increase from two fillers to three fillers is greater than an increase from five to six, which in turn will have more impact than an increase from seven to eight fillers.

Some jurisdictions in the U.S. use more than six. Some jurisdictions in Australia use lineups that vary from eight to 10 members, and England and Wales also use more than six members. Of course, as noted in the previous discussion on selecting fillers for lineups, it is not the nominal size of the lineup that matters so much as the number of lineup members who fit the description of the culprit. A lineup of 12 people would be less effective than a lineup of six people if the 11 fillers in the 12-person lineup did not fit the description of the suspect, whereas the five fillers in the six-person lineup did fit the description. Of course, a 12-person lineup in which all fillers were a good match to the suspect would provide more protection for the innocent suspect than would a six-person lineup in which all the fillers were good fillers. Although one eyewitness identification researcher has argued strongly for large increases in photo lineup sizes to as high as 120 (Levi, 2011), at this point we are not convinced that large increases in lineup size are warranted in practice. Among other things, all but one of the lineup members must be a priori cleared as possible suspects so that they can have the definitive status of known-innocent fillers. Establishing the innocence of a large number of fillers is not an issue in a lab experiment, but it would be in actual practice. As lineup size increases, it is also increasingly difficult to locate fillers who properly fit the description of the culprit. Adding extremely poor fillers to a lineup can enhance eyewitnesses' confidence in a mistaken identification (Charman, Wells, & Joy, 2011). In addition, there are concerns about potential loss of correct identifications (of the culprit) if a lineup becomes too large.

A Theoretical Note About Lineup Fillers

There is currently debate in the eyewitness identification literature about how good lineup fillers manage to improve overall lineup performance (i.e., how do they reduce mistaken identifications of innocent suspects more than they interfere with identifications of the culprit?). Some have suggested that the use of good fillers helps witnesses decide which facial features are diagnostic (e.g., Colloff et al., 2016; Wixted & Mickes, 2014), whereas others have argued that good fillers simply siphon false positive identifications away from the innocent suspect more than they siphon from the culprit (e.g., Smith et al., 2017; Smith, Wells, Smalarz, & Lampinen, 2018; Wells, Smith, & Smalarz, 2015). It is possible that both of these processes are involved and these might not be the only two possibilities. Regardless of which processes underlie the contribution that good fillers make to improved witness accuracy, the answer to this theoretical question ultimately could help to clarify the best strategies for choosing fillers for lineups, and should be considered in future research.

With photo lineups, care should be taken not only with respect to the choice of filler photos but also the photo of the suspect. Often, there are multiple possible photos of the suspect to choose from. When possible, the suspect's photo should be clear and look as much as possible like the suspect appeared at the time of the crime. If the crime was recent, for example, older photos of the suspect should not be used if a more recent photo is available.

Recommendation 5: Prelineup Instructions

When inviting an eyewitness to attend a lineup procedure (photo lineup or live lineup), police should not inform the eyewitness of any information that the witness has not already provided and certainly should not suggest that the suspect who will be in the lineup has been arrested or that the culprit will be present in the identification procedure. The eyewitness should be instructed that (a) the lineup administrator does not know which person is the suspect and which persons are fillers; (b) the culprit might not be in the lineup at all, so the correct answer might be "not present" or "none of these"; (c) if they feel unable to make a decision they have the option of responding "don't know"; (d) after making a decision they will be asked to state how confident they are in that decision; and (e) the investigation will continue even if no identification is made.

This recommendation addresses the concern that, if witnesses approach the identification test with the mistaken belief that the culprit must be present in the lineup, they may be predisposed toward making a positive identification. Intuition would suggest that many witnesses are likely to presume—based on the invitation to view a lineup—that the police must have a strong suspect and, thus, their task is to determine which lineup member the suspect is. There are various strands of evidence suggesting that witnesses make this assumption. For example, in one study, 90% of a large sample of witnesses indicated immediately after making their identification decision that they had expected the culprit to be present in the lineup and believed their task was to identify him or her (Memon, Gabbert, & Hope, 2004). In another study, witnesses either viewed a lineup containing the culprit or a lineup with the culprit removed but not replaced by another filler (Wells, 1993). In the former condition, 54% of witnesses picked the culprit, 25% picked one of the fillers, and 21% made no choice. When the culprit was removed but not replaced, it might be expected that around 75% of witnesses (i.e., 54% + 21%) would make no choice. Instead, only 32% of witnesses made no choice, with 68% distributing their choices across the various fillers. This finding suggests that witnesses are predisposed toward making a positive identification, though not necessarily with high confidence, provided some lineup member appears to be a reasonable match to their memory.

The most compelling evidence of the usefulness of this recommended instruction comes from studies comparing the identification performance obtained when witnesses are instructed that the culprit *might* or *might not be present* in the lineup to identification performance obtained under conditions where no such warning is provided. The former condition has typically been referred to as an unbiased instructions condition, the latter as a biased instructions condition. The way in which these two conditions have been enacted has varied. For example, unbiased instructions have often simply involved providing *the culprit might or might not be present* warning, although sometimes this has been accompanied by an instruction that there is no need to pick anyone or even an instruction emphasizing that the consequences of a wrong decision may be dire. Biased instructions may also take many forms, such as failing to forewarn the witness that the culprit might not be in the lineup or by strongly implying or even stating that the culprit is in the lineup. For example, asking the witness to select which lineup member is the culprit strongly implies that the witness is expected to make a positive identification decision.

Several features of the empirical findings on the effects of biased versus unbiased instructions warrant mention. First, the findings of three major reviews using meta-analytic procedures demonstrate that witnesses were more likely to make a positive identification decision when the lineup instructions were biased (i.e., no warning regarding possible absence of the culprit) than when they were unbiased (Clark, 2005; Steblay, 1997, 2013). Second, although the magnitude of this effect varied across the reviewed studies, presumably depending on the conditions at memory encoding and the identification test, the increased likelihood of choosing was reflected in increased positive identifications from both culprit-absent and culprit-present lineups. Thus, it had both positive and negative effects. Third, subsequent to the publication of these reviews, findings from two studies with very large sample sizes have reinforced the impact of biased versus unbiased instructions with both adult and child witnesses (Brewer & Wells, 2006; Keast, Brewer, & Wells, 2007). Biased instructions contributed to higher rates of both mistaken identifications and correct identifications, suggesting that a failure to warn witnesses that the culprit may not be present in the lineup contributes to witnesses being prepared to accept less evidence (i.e., establishing a more liberal decision criterion) for making a positive identification decision. This lower threshold for choosing could be beneficial if the base rate of guilty suspects were high, but it would be detrimental if the base rate of guilty suspects were low, as may often be true of real police lineups (e.g., Wixted et al., 2016).

The precise impact of witnesses lowering their decision criterion will also depend on factors such as the characteristics of the various lineup members (Brewer, Weber, & Semmler, 2005). For example, if the lineup is biased against the suspect by virtue of the suspect being the only plausible lineup member, a lower decision threshold would increase the likelihood of (a) a correct identification if the suspect is the culprit and (b) a mistaken identification of an innocent suspect if the suspect is innocent. In contrast, if the suspect is presented in an array of highly plausible fillers, instructions that lead witnesses to set a lower decision threshold may lead to responses being spread more evenly across all lineup members.

It is important to note one qualification of the pattern of findings typically found when contrasting the impact of unbiased versus biased instructions. The effect of unbiased instructions may be negated if the witness receives an explicit suggestion prior to viewing the lineup that the culprit may be present in the lineup. In a study in which witnesses were presented with culprit-absent lineups only, witnesses received the suggestion *surely you are going to be able to pick the person out from the lineup* prior to receiving the instruction that the culprit may or may not be present in the lineup (Quinlivan et al., 2012). Witnesses who received that

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

suggestion, followed by unbiased instructions, were three times more likely to identify the suspect (who, in this study, was innocent) than were witnesses who received unbiased instructions without any prior suggestive comment. Witnesses in the former condition were also more confident in the accuracy of their erroneous identification decision than those in the latter condition.

The research findings on instructing witnesses prior to their viewing a lineup have clear implications. First, when inviting an eyewitness to attend a lineup procedure, police should not suggest that a suspect has been arrested or that the culprit will be present in the identification procedure. Second, in our experience some witnesses seem to be under the misconception that the investigation hinges on their identification decision. Consequently, witnesses should also be told that the investigation will continue even if no identification is made.

Third, it should be made quite clear to the witness that the culprit may or may not be in the lineup and that they do not have to select any of the lineup members. In other words, responses such as *not present* or *none of these* are quite appropriate. A reminder that the witness does not have to choose anyone from the lineup is important. A large percentage of witnesses are under the impression that the culprit is present and their task is to identify him (Memon et al., 2004). Fourth, to ensure that the witness does not lose sight of the fact that such response options are appropriate, there should be an explicit *not present* response option accompanying the lineup members from which the eyewitness can choose. In the case of a photo lineup, this option may be located below the array of lineup faces. In the case of a live lineup, a response sheet that shows the possible response options can be used: lineup member numbers (i.e., 1, 2, . . . , 6), not present, and do not know.

Finally, lineup administration procedures should accommodate the possibility that the witness may look at the lineup and be unwilling to pick someone or to respond not present because, for example, they cannot decide between two or more lineup members or they are uncertain about whether the culprit is in the lineup. For that witness (i.e., one who really has no idea about what to do), an appropriate response may be to say "don't know" rather than not present. Both adult and child witnesses use options such as do not know or not sure when they are made explicitly available, with frequency of use varying considerably depending on the encoding stimuli and lineup materials (Brewer, Keast, & Sauer, 2010; Weber & Perfect, 2012; Zajac & Karageorge, 2009). Moreover, there is some evidence indicating that positive identifications of a suspect are more diagnostic of suspect guilt when they are made in the presence of a do not know option compared with when no such option existed (Weber & Perfect, 2012).

The availability of an option to respond do not know is likely to reduce the likelihood of low confidence positive identifications, which research shows are often inaccurate (Brewer & Wells, 2006; Wixted & Wells, 2017). Although there are strong grounds for always questioning the reliability of low confidence identifications, there may be a tendency on the part of police or prosecutors to argue, for example, that the initial low confidence identification was made from a photo that was not a good likeness to the suspect, thereby purportedly providing a reasonable explanation for the witness's low confidence. Further, although studies have shown that mock-jurors (appropriately) downgrade the credibility of the witness and the culpability of the defendant when cross-examination highlights a disparity between a witness's expressed confidence at initial identification (low confidence) versus an in-court identification (high confidence; Bradfield & McQuiston, 2004; Jones, Williams, & Brewer, 2008), this effect does not always occur (e.g., Douglass & Jones, 2013). For example, if a witness expressed high confidence in their identification of the suspect during a trial despite a low confidence initial identification, jurors tended to excuse the confidence inflation if the witness appeared to have experienced an epiphany about their initial low confidence identification (e.g., "I wasn't very confident at the time of the identification because I was scared back then"). As long as these types of excuses for initial low-confidence identifications are permitted, we argue that it is crucial that there is an explicit do not know response option, which can be located alongside the not present option.

We note that some jurisdictions have used what has been referred to as an "appearance change" instruction. This instruction was among a set of guidelines developed by a U.S. Department of Justice (DoJ) working group on the collection of eyewitness evidence. Specifically, the DoJ guidelines recommended that, prior to being shown a lineup, eyewitnesses should be told that "individuals depicted in lineup photos may not appear exactly as they did on the date of the incident because features such as head and facial hair are subject to change" (Technical Working Group for Eyewitness Evidence, 1999, p. 32). We have not included this instruction in the current set of recommendations because subsequent research has shown that the appearance change instruction increased false identifications but did not increase culprit identifications (Charman & Wells, 2007; Molinaro, Arndorfer, & Charman, 2013).

As a final note on instructions, we recommend that in addition to the witness having the instructions in writing, the lineup administrator should read the instructions aloud to witnesses, pausing after each point to make sure that the witness understands each point.

Recommendation 6: Obtain an Immediate Confidence Statement

A confidence statement should be taken from witnesses as soon as an identification decision (either positive or negative) is made.

For double-blind lineups, "immediate" means that the confidence statement should be secured with only the blind administrator in the room and before the case detective or any other nonblind individuals are allowed into the room. Note that a confidence statement should be recorded if a witness positively identifies someone or if a witness indicates the culprit is *not present*, says they *do not know*, or indicates that they are *not sure*. As an example, the confidence statement could be collected as a numeric response (i.e., on a scale from 0% confident to 100% confident). Alternatively, confidence could be collected using a verbal scale (e.g., "positive," "probably," "maybe"). If neither scale is used and witnesses simply use their own words, a verbatim record of their verbal statements (or preferably a video-recording, see Recommendation 7) should be made, not a summary or paraphrase generated by the lineup administrator. If the witness's response is "don't know," a confidence statement should be recorded if the witness spontaneously provides one. Otherwise, no confidence statement should be solicited for "don't know" responses. However, it could be useful to let the eyewitness state a basis for the "don't know" response (e.g., did not get a good view, none are familiar). The prelineup instructions should have already communicated to the eyewitness that a confidence statement will be requested (see Recommendation 5). These prelineup instructions help to prevent witnesses from drawing erroneous conclusions that their confidence is only being assessed because the lineup administrator thinks their decision is incorrect.

This recommendation is based on the fact that eyewitness confidence is a useful cue to the accuracy of a witness's decision when instructions do not imply the presence of a culprit (Quinlivan et al., 2012), double-blind administration is used, and fair lineups are presented (Wixted & Wells, 2017). Specifically, confidence predicts accuracy among witnesses who choose from a photospread when immediate confidence reports are obtained (e.g., Brewer & Wells, 2006; Palmer, Brewer, Weber, & Nagesh, 2013; Sauer, Brewer, Zweck, & Weber, 2010). Note that forcing witnesses to withhold their confidence reports for as little as 5 min has been found to undermine the predictive value of confidence, unless witnesses are required to spend that time reflecting on the witnessing conditions and identification procedure, or producing reasons why they might have made an incorrect decision (Brewer, Keast, & Rishworth, 2002).

The corpus of data suggesting that confidence is a useful cue to eyewitness accuracy stands in stark contrast to DNA exonerations in which innocent people were mistakenly identified in court by highly confident eyewitnesses, most of whom were demonstrably less confident at the time of the initial identification (e.g., Garrett, 2011). If confidence is a useful cue to accuracy, how could these mistaken witnesses have been so confident? The answer lies in the fact that intervening postdecision events can dramatically shift witnesses' reports of their confidence from low at the time of identification to high at the time of trial. As a result, these shifts render delayed reports nearly useless as cues to accuracy. However, if confidence reports are taken immediately after an identification decision, the integrity of confidence as a cue to accuracy is enhanced considerably.

We recommend that confidence statements be collected on a graded scale using words (e.g., "positive," "probably," "maybe") or numbers (e.g., from 0% confident to 100% confident). The key element of this recommendation is that an immediate record of a witness's confidence is collected. Immediate confidence estimates are the only way to ensure that postidentification variables do not contaminate subsequent confidence reports.

We focus on *immediate* confidence reports because seemingly innocuous postdecision events can contaminate witnesses' confidence reports, undermining what could have been forensically meaningful information from an eyewitness. One of the most heavily researched of these events involves postidentification feedback offered by a lineup administrator in the form of a simple comment confirming the witness's decision. Such feedback can dramatically inflate confidence reports. In the original test of the postidentification feedback effect (Wells & Bradfield, 1998), researchers provided inaccurate witnesses with such a comment: "Good, you identified the suspect." That simple statement resulted in 50% of inaccurate witnesses reporting that their confidence was a 6 or 7 on a 7-point scale (compared with only 15% of witnesses in the control condition). Importantly, the inflated confidence report created by this feedback is a *retrospective judgment* because witnesses indicate how confident they were at the time of their identification, before they knew their decision was correct. Any resulting confidence inflation obviously obscures a true picture of the witness's experience at the time of the identification decision. The effect of postidentification feedback is robust and reliable (see meta-analysis by Steblay et al., 2014). It also features prominently in some judicial decisions as courts grapple with how to ensure that eyewitness identification testimony truly reflects the witness's experience of making an identification decision, rather than the influence of extramemorial variables (e.g., *New Jersey v. Henderson*, 2011; *Oregon v. Lawson*, 2012).

Beyond affecting witnesses' self-reports, distorted confidence judgments complicate assessments of witness identification decisions. Indeed, several experiments show us that evaluators rate inaccurate witnesses who have received confirming feedback as more credible than those who received disconfirming feedback or no feedback, even when evaluators are instructed to ignore the feedback and even when an explicit confidence statement is not available (Douglass, Neuschatz, Imrich, & Wilkinson, 2010). Because postidentification feedback inflates the confidence of inaccurate witnesses more than the confidence of accurate witnesses, it also impairs fact finders abilities to distinguish accurate from inaccurate witnesses (Bradfield et al., 2002; Smalarz & Wells, 2014b).

Recommending an immediate confidence report is an important companion to the recommendation that the entire lineup procedure be video-recorded (see Recommendation 7). Not only can video-recording demonstrate whether postidentification feedback was given, but it can also preserve witness nonverbal cues that may signal accuracy (Matuku, Douglass, & Charman, 2018). In the absence of a video-recorded identification procedure, triers of fact who learn that a witness's confidence has inflated over time are sometimes unwilling to impugn the witness's credibility (Bradfield & McQuiston, 2004), especially if a compelling explanation accompanies the inflation (e.g., "I had an epiphany!"; Jones et al., 2008). However, if evaluators *see* the identification procedure in which a witness's initial confidence is lackluster, their assessments of a highly confident trial witness are less positive, which is an important shift when the highly confident witness has identified the wrong person (Douglass & Jones, 2013).

Beyond distorting evaluators' ability to assess witnesses, contaminated witness reports may also unduly shape preliminary investigations by (a) triggering biased evaluations of subsequent pieces of evidence and/or (b) biasing the integration of evidence against the identified suspect. For example, a witness who is highly confident in a mistaken identification may trigger investigators to view the suspect's alibi as weaker than it would have been otherwise or may suggest to investigators that they suspend pursuit of additional suspects (for a discussion of these effects see Charman, Douglass, & Mook, 2019). Recording a witness's immediate confidence eliminates the potential for subsequent inflations to go unnoticed.

If lineup administrators follow the recommendation to conduct double-blind procedures (Recommendation 3), they will be unable to provide postidentification feedback because they will not know which person is the suspect. Therefore, they will be unable to confirm (or disconfirm) any decision made by the witness. However, even vague positive comments ("You have been a good witness") can be interpreted by witnesses as confirming feedback (Dysart, Lawson, & Rainey, 2012). Therefore, consistent with the

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

recommendation on prelineup instructions (Recommendation 5), witnesses should be explicitly told that the lineup administrator does not know which person is the suspect and which people are fillers.

Even if administrators correctly implement double-blind procedures and prelineup instructions, it is still imperative to collect immediate confidence reports because witnesses may encounter postidentification feedback in multiple other ways. For example, the mere fact that a case proceeds to trial is a form of confirmation that a prior identification was correct (see also Berkowitz & Frenda, 2018; Wade, Nash, & Lindsay, 2018 for other types of real world contamination). Another potential source of contamination comes from witnesses who conduct their own investigations via social media searches, often accompanied by cowitnesses who may be able to "confirm" that an Instagram or Facebook photo represents the culprit (Douglass & Smalarz, 2019).

Recommendation 7: Video-Recording

The entire identification procedure, including prelineup instructions and witness confidence statement, should be video-recorded.

To preserve a faithful record of the conditions under which witnesses make their identifications, we recommend that the entire interaction between the police and the witness be video-recorded—from the time when the witness is given the very first prelineup instructions through to the completion of the procedure when the witness has provided identification and confidence statements. The video must capture all features of the administration, including the interactions among the lineup administrator, the witness, and the lineup members (whether they are presented live or by photo). Under some circumstances, fully capturing the lineup administration may necessitate cameras recording from multiple angles. Although the authors of the original scientific review paper (Wells et al., 1998) recognized some of the benefits of video-recording a lineup administration, they did not include video-recording as one of their recommendations at that time. In part, video-recording the identification procedure was left out of the original set of recommendations because of concerns about the costs associated with making the record, including costs for equipment and materials. Since that time, the cost of video-recording interactions has decreased considerably and most adults have cellular phones capable of rendering high-quality video-records. Furthermore, increasing numbers of police now have access to body cameras that can be positioned to make video-recordings of identification procedures. Thus, we believe it is time to make the video-recording of lineup administrations standard practice as it is in some jurisdictions and in other countries (e.g., Australia).

In his response to the initial scientific review paper, Kassin (1998) noted two reasons why video-recording identification procedures should be considered best practice. First, police reports of what happened during an identification procedure may be incomplete or even inaccurate given that they are based on officers' recollections of what happened during the procedure; recollections that are subject to the typical foibles of human memory. Second, it is possible that video-recording the identification procedure could encourage administrators to carefully adhere to best practices and deter them from engaging in any suggestive practices (Kassin, 1998). In the remainder of this section, we expand upon each of these justifications for video-recording identification procedures, citing relevant research when it is available.

Clearly, video-recording identification procedures has the benefit of providing a more precise and accurate accounting of what happened during the procedure. Although it might be a rare occurrence, video-recording the procedure will make it more difficult for police officers to intentionally fabricate their reports of what occurred during the lineup administration. There is evidence of some police officers purposefully misrepresenting case-related events (Orfield, 1987, 1992; Slobogin, 1996), but even in the case of a conscientious, well-meaning officer, there are benefits of a more accurate reporting of the procedure through video-recording. For example, the memories of police officers are subject to the same cognitive errors as are those of others, including interference (e.g., Kane & Engle, 2000; Nairne, 2002) and memory intrusions from mental scripts about what usually occurs (e.g., Greenberg, Westcott, & Bailey, 1998; Kleider, Pezdek, Goldinger, & Kirk, 2008). One can easily imagine that memories from other lineups conducted or mental scripts for what should have happened could interfere with what a police officer remembers, and consequently reports, about an identification procedure. Indeed, there have been cases in which a police officer testified to reading mandated instructions verbatim to the witness, whereas the video-recording subsequently revealed improvisation that introduced suggestiveness into the procedure. Whether or not the officer was attempting to intentionally mislead or was genuinely mistaken, the actual procedure could be easily reviewed if it is video-recorded, and the fallibility of memory is no longer an issue.

In contrast, witnesses' memories for a procedure are not likely to be influenced by interference or mental scripts regarding what usually happens given that most witnesses have never participated in a lineup before. However, their lack of expertise with the procedure will likely result in reports that are less complete (e.g., Chase & Simon, 1973; Gobet & Simon, 1996). Moreover, their reports of the procedure may be affected by the decay of their memory trace over time (Deffenbacher, Bornstein, McGorty, & Penrod, 2008) or postevent information (Loftus, 2004). The video-record would also provide clear information about how long it took the witness to make an identification. Witnesses' estimates of time are often inaccurate (Yarmey, 2000), but time-to-identification is an important postdictor of witness accuracy: Identifications made more quickly are more likely to be accurate than are those made more slowly (Sporer, 1992). Although research has not provided a definitive cutoff that allows us to discriminate between accurate and inaccurate identifications (Brewer, Caon, Todd, & Weber, 2006; Weber, Brewer, Wells, Semmler, & Keast, 2004), having an accurate record of a witness's time-to-identification can provide useful information to fact finders. In sum, despite knowing no evidence directly examining the accuracy of police officers' and witnesses' memories of identification procedures, we are confident that the basic cognitive research on memory errors generalizes to this context.

Although there is no direct evidence about the accuracy of police reports of identifications, police reports of witness interviews and suspect interrogation procedures omit important details about the procedures used (Kassin et al., 2017; Lamb et al., 2000). Moreover, testimony from witnesses about what happened during the procedure is even more likely to be subject to error given that police reports are likely to be written shortly after the event,

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

whereas testimony may be given months or years later. Video-recording the identification procedure allows police command staff to confirm that proper procedures are being followed. In addition, video-recording the procedure eliminates the need to rely on police officers' or witnesses' memories for the event at trial. In many cases, detectives might not have to testify during suppression hearings if there is a video available, thereby saving staff resources and money by not having to pay officers to attend court.

When deciding a motion to suppress the identification, the judge could review the video and evaluate the suggestiveness of the procedure herself rather than relying on attorneys' characterizations of the procedure based on their readings of police reports and witness testimony. Defense attorneys might review the video-recording when deciding whether to offer a plea or encourage a client to accept one, whereas prosecutors might review the video-recording when evaluating the strength of a case and how to proceed. Moreover, in the United States a suspect does not have the right to have an attorney present at a preindictment identification procedure and never has the right to have an attorney present at a photo lineup procedure (*United States v. Ash*, 1973). In the absence of an attorney to view the identification procedure, the video-recording could help defense attorneys to better represent their clients following identification. In addition, eyewitness experts could be asked to review the video-recording and either testify or prepare a written report about the identification procedure. Finally, the video-recording could be introduced as evidence at trial so that fact finders can judge for themselves whether the procedure was suggestive, whether the witnesses engaged in any behaviors that either enhanced or diminished their credibility, or whether the witness's confidence was inflated (e.g., Douglass & Jones, 2013).

Although the data on this topic are limited, and there are no known data on how judges or attorneys might evaluate video-records of identification procedures, there are a handful of studies that explore the effects of watching the video-recording of an eyewitness identification procedure on mock juror judgments. Most of these studies have tested whether viewing a video-recording of the identification procedure helps jurors evaluate the accuracy of witness identifications (Beaudry et al., 2015; Reardon & Fisher, 2011). In one study, watching the video-recording helped participants distinguish between accurate and inaccurate witnesses (Reardon & Fisher, 2011). In another study (Beaudry et al., 2015), confirmatory feedback interfered with the ability of participant-judges to differentiate between accurate and inaccurate witnesses, with participants judging the witnesses who received confirmatory feedback to be more accurate irrespective of their actual accuracy. Participants' judgments of accuracy were unaffected by whether the lineup was conducted using single- or double-blind procedures, irrespective of whether the participants had viewed the video of the procedure. However, these videos were relatively short (M = 1 m, 37 s) and it is not clear from the report of the study how much influence the administrators exerted in the single-blind conditions. In a study that manipulated whether the videotape contained evidence of administrator influence, watching a video-recording of a single-blind versus a double-blind procedure did influence jurors' verdicts (Modjadidi & Kovera, 2018). Specifically, watching the video-recording increased participants' ratings of procedural suggestiveness and decreased their guilty verdicts when the video-recording depicted witness steering that can occur in a single-blind administration as opposed to a double-blind administration. Thus, the limited evidence available suggests that watching the video-recording of a lineup administration will help jurors (and perhaps judges and attorneys) evaluate whether a lineup procedure was suggestive.

Finally, video-recording the identification procedure could encourage administrators to ensure that their conduct conforms to best practices and deter them from engaging in any suggestive procedures. Although we know of no studies that directly test whether video-recording can change the behaviors of administrators of identification procedures, the issue has been examined in the context of interrogation practices. In a mock crime and interrogation paradigm, researchers manipulated the actual guilt-innocence of suspects who were then subjected to interrogations conducted by police officers who were either informed or uninformed regarding the researcher's recording of the interrogation session (Kassin, Kukucka, Lawson, & DeCarlo, 2014). Police officers who knew that they were being recorded were significantly less likely to use minimization tactics and somewhat less likely (although not significantly so) to use maximization tactics; both tactics are known to increase rates of false confessions (Kassin et al., 2010). In addition, participant-suspects (who were made aware of the video-recording manipulation) reported that the police officers who did not know they were being recorded tried harder to make them confess than did police officers who knew about the recording. Thus, there is evidence, albeit from outside the eyewitness arena, that video-recording police procedure can deter undesirable behaviors.

There are few data on the question of whether video-recording eyewitness identification procedures might change the behavior of eyewitnesses. But there are good data indicating that even crime suspects are not inhibited by video-recording during interrogations and that people quickly forget that they are being recorded (Kassin et al., 2019).

In sum, both logic and the available empirical evidence support the recommendation to video-record all identification procedures. If the procedure is to be recorded, it should be recorded in a way that captures all relevant information about the procedure, including the verbal and nonverbal behaviors of the witness, the administrator, and the lineup members. Research from the interrogation literature makes it clear that camera angle matters, in that people attribute causality to the person who is the focus of the video-recording (Lassiter, 2010; Lassiter et al., 2009; Ratcliff et al., 2006). In the case of interrogations, a camera focused on the suspect causes evaluators to be more likely to view a confession as voluntary and more likely to judge them to be guilty than if the camera was focused on the interrogator or equally on the suspect and the interrogator. It is reasonable to believe that a camera focused on the witness may similarly cause evaluators to overlook suggestive behaviors on the part of administrator, or features and behaviors of suspects that make them stand out from the other lineup members. However, a video-recording of an identification procedure that includes information about the witness, the administrator, and lineup members can provide a complete record of the procedure that documents suggestive practices when they are present and protects the police from unjustified and time-consuming claims of bias when the procedure was free from suggestion.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

This document is copyrighted by the American Psychological Association or one of its allied publishers. Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

Recommendation 8: Avoid Repeated Identification Procedures With the Same Witness and Suspect

Repeating an identification procedure with the same suspect and same eyewitness should be avoided regardless of whether the eyewitness identified the suspect in the initial identification procedure.

This recommendation holds no matter how compelling the argument in favor of a second identification might seem (e.g., the original photo of the suspect was not as good as it could have been; the witness was nervous during the first identification test and is calmer now; the initial identification was made from a social media profile, but it would be more desirable to have an identification made using proper police procedures). The importance of focusing on the first identification test cannot be emphasized strongly enough.

There are certain kinds of forensic evidence for which repeated testing is not only acceptable but also desirable. A crime scene fingerprint comparison with a suspect's prints, for example, might be subjected to repeated comparisons to confirm a conclusion. Likewise, given a sufficient amount biological material from a crime scene, forensic examiners might test only a small portion of the sample for DNA, allowing for subsequent testing by a different analyst or different means of evaluation. However, eyewitness identification evidence has a unique characteristic that makes it unsuitable for what might be called "repeated testing." Whether the eyewitness is asked to make an identification with a showup or a lineup, there is only one *uncontaminated* opportunity for a given eyewitness to make an identification of a particular suspect. Any subsequent identification test with that same eyewitness and that same suspect is contaminated by the eyewitness's experience on the initial test.

For purposes of our recommendation, repeated identification tests refer to a situation in which an eyewitness is given a subsequent identification test (or more) with the *same* suspect that appeared in an earlier identification test. We are not referring to a situation in which an eyewitness is given an identification test with one suspect and then, after rejecting that suspect, is given an identification test with a *different* suspect and different fillers. Nor are we referring to a situation in which there were multiple culprits and the eyewitness is given a separate identification test for each culprit.

There are many ways in which the use of repeated identification procedures surface in criminal cases. For example, eyewitnesses might view a mug book that contains the suspect prior to viewing a lineup that includes that suspect, or an eyewitness might first view a showup and then later be shown a lineup with that same suspect. A repeated identification can also occur when the eyewitness views a lineup and makes a tentative pick (or no pick) and then, at a later time, sees a second lineup with that same suspect and new fillers. Sometimes a witness is first shown a photo lineup and then later is shown a live lineup using that same suspect. Other times the first lineup uses one photo of the suspect, and the second lineup contains a different photo of the same suspect. Perhaps the most common repeated identification procedure of all is when the witness makes an out-of-court identification (from a showup or a lineup) and then is asked to repeat that identification in court (i.e., an in-court or "dock" identification) at pretrial hearings or at trial or at both the pretrial hearing and at trial.

At a theoretical level, there are at least three processes by which an initial identification test that includes a given suspect can *contaminate* a later identification test if the later test includes that same suspect (Deffenbacher, Bornstein, & Penrod, 2006; Steblay & Dysart, 2016). One such process is memory-source error (or "source monitoring error"; see Johnson, Hashtroudi, & Lindsay, 1993). Memory-source error is implicated when the eyewitness perceives the suspect in the second identification procedure to be familiar and misattributes the familiarity to the original witnessed event rather than to the fact that the previous identification procedure included that face. In this type of effect, even though the eyewitness did not identify the suspect in the first identification procedure, that person's face was made familiar by its appearance in the first procedure. This source misattribution effect, which involves a dissociation between familiarity and an awareness of the source of that familiarity, was first demonstrated over 40 years ago (Brown, Deffenbacher, & Sturgill, 1977). It is closely related to other phenomena such as familiarity without awareness (Mandler, 1980), the false fame effect (Jacoby, Woloshyn, & Kelley, 1989), and the "bystander effect," in which a bystander to the crime is identified as the culprit due to a misattribution of familiarity (Ross, Ceci, Dunning, & Toglia, 1994). Simply being exposed to an innocent suspect in a mug book, showup, or a lineup increases the chances of that person being identified in a later lineup even if the witness did not choose the person in the first identification procedure (e.g., Brown et al., 1977; Dysart, Lindsay, Hammond, & Dupuis, 2001; Haw, Dickinson, & Meissner, 2007; Hinz & Pezdek, 2001; Lawson & Dysart, 2014; Steblay, Tix, & Benson, 2013).

A second process by which the first identification procedure can contaminate the later identification procedure is when the eyewitness identifies the suspect in the initial identification procedure and is later given another identification procedure with that same suspect and a different set of fillers. In this case, the initial identification, even if mistaken, causes the witness to simply repeat the same identification in the second identification procedure. A meta-analysis of this *commitment effect* has provided strong evidence that a mistaken identification in an initial identification procedure tends to be repeated in a second identification procedure if that lineup contains the mistakenly identified person (Deffenbacher et al., 2006). More recent studies have provided additional support for this effect (e.g., Godfrey & Clark, 2010; Goodsell, Neuschatz, & Gronlund, 2009; Hinz & Pezdek, 2001; Lawson & Dysart, 2014; Pezdek & Blandon-Gitlin, 2005; Steblay et al., 2013; Valentine, Davis, Memon, & Roberts, 2012). Researchers have argued that commitment, which is the powerful tendency to stick with an earlier decision that was freely made, is the psychological mechanism underlying this effect. But commitment is not the only process by which an initial mistaken identification results in a repeat mistaken identification of the same person in a subsequent identification test. For example, there is evidence that the act of identifying an innocent person in an initial identification procedure changes the eyewitness's memory away from the culprit and toward the person identified, a process that is intensified if the witness is given confirming feedback following the initial mistaken identification (Smalarz & Wells, 2014a).

Although not yet specifically tested in controlled experiments, there is a third possible process by which repeating the same suspect in a second identification procedure can contaminate the second identification decision. Specifically, it is possible for the eyewitness who makes no identification in the first procedure (e.g., a photo lineup) to later consciously and explicitly recognize that there is only one person in the second procedure (e.g., a live lineup) who was also in the first procedure. In effect, this "gives away" the hypothesis of police investigators by communicating to the eyewitness which person is the suspect (i.e., the person in common between the two procedures). In other words, this type of repeated procedure violates a fundamental characteristic of a good lineup, namely that there not be aspects of the procedure that leak information about which person is the suspect in the lineup (Wells & Luus, 1990).

Any of these processes might occur in a given situation that involves repeating the suspect in an identification procedure. For this reason, we recommend that repeated identification procedures be avoided. Of course, it could be argued that if the witness identifies the suspect in the first identification procedure then there is no harm in repeating the identification procedure. After all, the identification has already been made and the second identification procedure is merely a confirmatory process. However, there is good evidence that repeated testing of eyewitnesses leads to artificially elevated levels of eyewitness confidence (Shaw, Garven, & Wood, 1997; Shaw & McClure, 1996). Repeatedly asking a question appears to lead to increased ease or fluency of answering the question, which leads witnesses to develop a stronger sense of confidence that their answer is correct, even when their answer is incorrect (Shaw, 1996).

Finally, it is important to note that when witnesses make identifications outside of a police procedure, any additional identification procedure conducted by the police is a second identification attempt. Witnesses may spontaneously identify someone as the culprit as they walk about their daily lives. Sometimes witnesses to crimes launch their own investigations. They may hear the culprit referred to by a nickname or street name and then search social media for someone using that name. Or they may search the social media connections of someone they knew was present during a crime looking for the culprit. Whether an identification is made through a self-directed search of social media or spontaneously spotting a suspect on the street, this identification is the first identification, and it will contaminate any subsequent identification procedure the police might present to the witness.

Recommendation 9: Showups

Showups should be avoided whenever it is possible to conduct a lineup (e.g., if probable cause exists to arrest the person then a showup should not be conducted.). Cases in which it is necessary to conduct a showup should use the procedural safeguards that are recommended for lineups, including the elimination of suggestive cues, a warning that the detained person might not be the culprit, video-recording the procedure, and securing a confidence statement.

The term *showup* refers to an eyewitness identification procedure in which a single individual is presented to the eyewitness and the eyewitness is asked whether this is the person who committed the crime in question. The primary defining feature that separates a showup from a lineup is the use of fillers: A showup has no fillers, whereas a lineup does. Showups have been heavily criticized as being extremely suggestive, a criticism that dates back more than 100 years (Gross, 1911). In 1967, the U.S. Supreme Court, in reference to a showup identification, said that "It is hard to imagine a situation more clearly conveying the suggestion to the witness that the one presented is believed to be guilty by the police" (*United States v. Wade*, 1967, p. 226). Despite the strong language seemingly condemning showups, the U.S. Supreme Court has consistently supported the admissibility of showups (e.g., see *Manson v. Braithwaite*, 1977; *Neil v. Biggers*, 1972) provided that the witness shows evidence of reliability in other ways. The Court listed five factors to consider that might indicate that a suggestive procedure could nevertheless be reliable, namely that (a) the witness had a good opportunity to view the culprit, (b) the witness paid good attention while witnessing, (c) the witness gave a good description of the culprit, (d) there was a short delay between the witnessed event and the identification, and (e) the witness made the identification with high confidence. These criteria have been strongly criticized by eyewitness scientists on a variety of counts, especially when the procedure was suggestive (Wells & Quinlivan, 2009).

In practice, showups tend to be conducted under specific conditions. In particular, showups tend to be restricted to situations in which an individual who fits the description of the culprit of a crime is detained in the general vicinity of the crime shortly after the crime has occurred. As discussed in the introduction to this article, there is a rationale for sometimes permitting showups under these conditions (fit description, proximity to crime, and soon after the crime occurred). The rationale is that there is not enough time to construct and conduct a lineup procedure because the detained person can only be detained for a relatively short period of time unless there are grounds for arrest. Fitting a description of the culprit is not, in and of itself, grounds for arrest. So, unless there is probable cause for arresting the detained person, the choice is not between conducting a lineup and conducting a showup; rather, the choice is between conducting a showup or setting the detained person free, thereby potentially creating a public safety issue. Moreover, because showups sometimes result in eyewitnesses rejecting the detained person as being the culprit (see Gonzalez, Ellsworth, & Pembroke, 1993), showups can result in innocent people being quickly excluded as possible suspects, thereby allowing a continuation of the search for the true offender.

About 15% of DNA exonerations involving eyewitness identification involved showups (West & Meterko, 2017). Experiments comparing lineups with showups reveal that lineups are clearly superior to showups in terms of the lineup procedure's ability to distinguish between innocent and guilty suspects (e.g., Clark, 2012a; Dekle, Beal, Elliot, & Huneycutt, 1996; Gronlund et al., 2012; Mickes, 2015; Steblay, Dysart, Fulero, & Lindsay, 2003; Wetmore et al., 2015; Yarmey, Yarmey, & Yarmey, 1996; see meta-analysis by Neuschatz et al., 2016). Interestingly, however, witnesses are no more likely to make an affirmative identification with a showup than with a lineup (Gonzalez et al., 1993; Wells, 2001). In fact, witnesses actually make *more* affirmative identification responses to lineups than to showups, perhaps because there are more people from whom to choose. However, inaccurate identifications from showups always fall on the innocent suspect, whereas in lineups such inaccurate choices tend to spread across known-innocent fillers (Smith et al., 2017; Wells, 2001). In several ways, the fillers used in a lineup act as a safeguard, protecting the innocent suspect from mistaken identification—a form of protec-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

tion that does not exist for showups (see Wixted & Mickes, 2014, for a different account of why lineups work better than showups).

Our recommendation is that showups should be avoided (and a lineup conducted instead), if at all possible. For example, a showup can be avoided and a lineup used instead if there is probable cause for arresting a detained person instead of doing a showup. If, for example, the detained person is in possession of stolen goods, is in possession of an unauthorized firearm or other weapon, or has an outstanding warrant, then the person should be arrested. Indeed, without evidence linking the detained person to the crime (other than being in the vicinity of where it was committed, which—as noted previously—is not evidence-based suspicion), officers are using the most suggestive identification procedure (a showup) when base rates of guilt are the lowest. A lineup for the arrested individual can then be arranged rather than a showup. Another situation in which showups can sometimes be avoided occurs when there are multiple witnesses. In cases of multiple witnesses, a showup can be conducted with one of the witnesses and, if an identification is made, the identification becomes grounds for arrest. The remaining witnesses can then be preserved for more reliable lineup procedures, which should be conducted only after evidence is developed connecting the suspect to the specific witnessed crime.

Reducing the Suggestiveness of Showups

Although showups are clearly less reliable than lineups, there are some ways to reduce the suggestiveness of showups. First, many of the features of a good lineup procedure can be incorporated into showup procedures. For example, recent research finds that a preshowup instruction about additional opportunities can be effective in reducing mistaken identifications with little or no reduction in accurate identifications (Eisen, Smith, Olaguez, & Skerritt-Perta, 2017; Smith, Wells, Lindsay, & Myerson, 2018). This additional-opportunities instruction simply tells witnesses prior to viewing the detained person that if they do not think the person is the culprit, they might have additional opportunities to view someone else. The theory behind the additional-opportunities instruction is that witnesses set a low criterion for making an identification with showups because they assume that this is their only opportunity to identify the culprit.

As with lineups, eyewitnesses who participate in a showup procedure should be asked to report how confident they are in their identification or rejection decision. In addition, as with lineups, showups should be video-recorded, a recommendation that is relatively easy to implement now that dashboard-mounted and body-worn cameras are increasingly commonplace (see Recommendation 7). The video of the showup should start before the witness is instructed and continue through the witness's confidence statement, and should include the officer, witness, and suspect in view.

Because showups are conducted in the field during which search-and-detain operations are actively unfolding, care should be taken to ensure that witnesses do not overhear police radio conversations that could prejudice a showup procedure. In an experiment using a high-realism paradigm for studying showup identifications, overhearing the suggestion that the sheriff had caught the guy significantly increased false identifications from showups but did not affect accurate culprit identifications (Eisen, Skerrit-Perta, Jones, Owen, & Cedre, 2017).

Clothing is often a central feature of an eyewitness's description that police use for finding a person who is then detained for a showup. The concern is that the eyewitness might identify the person based primarily on the clothing (see Lindsay et al., 1987). In cases in which the person was detained based on a clothing description, consideration should be given to covering the person's clothing with a blanket for the showup identification test.

For legal reasons, eyewitnesses are usually brought to the location of the detained person for a showup rather than the detained person being brought to the eyewitness, because transporting the detained person is usually considered an arrest. Consistent with our recommendations, if the detained person can be arrested, then a lineup should be conducted rather than a showup. Taking the eyewitness to the detained individual rather than the reverse also makes it easier in multiple-witness cases to ensure that the witnesses are not contaminated by observing an identification decision made by another witness. As noted previously, if one witness makes an identification, then that would normally be considered probable cause for arrest and the remaining witnesses can later be shown a lineup instead of a showup. Placing the detained person in handcuffs or in the back seat of a patrol car for the showup can suggest to the witness that the person has been arrested. This indication that the person has been arrested, in turn, can suggest to the eyewitness that there is evidence against the detained person beyond simply matching the description of the suspect. Hence, unless there is reason to believe that the detained person is a flight risk, these types of restraints should be avoided during a showup.

Severe limits on safeguards with showups. Obviously, there are some critical safeguards that can be used with lineups that cannot be used with showups. By definition, showups do not include fillers, which is a key safeguard. But showups also cannot be double-blind. In fact, showups cannot ever be single blind. After all, both the showup administrator and the eyewitness know which person is the suspect, namely the person being presented to the eyewitness.

In-court identifications as showups. The current article concerns policies and procedures for collecting and preserving eyewitness identification evidence rather than how courts handle eyewitness evidence. Nevertheless, there is a common courtroom procedure known as an in-court identification (or a dock identification) in which an eyewitness on the stand is asked if she or he can identify the culprit in the courtroom. The defendant typically is sitting at the defense counsel table rather than embedded among known-innocent fillers who fit the description of the culprit. Hence, an in-court identification is closely analogous to a showup. In terms of suggestiveness, the in-court identification is arguably even more suggestive than a typical showup because it clear to the witness that the defendant has already been indicted. Moreover, it is usually the case that the eyewitness has already identified the defendant in a precourt procedure, which means that the in-court identification is a repeated identification that goes directly against Recommendation 8 (e.g., avoid repeated identifications). And, if the eyewitness has not already identified the defendant in a proper precourt procedure, the showup nature of an in-court identification should not be considered an acceptable alternative to a properly constructed and conducted lineup. The low probative value of an

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

in-court identification raises serious concerns that its prejudicial value exceeds it probative value.

Final Observations

The current set of recommendations for the collection and preservation of eyewitness identification evidence is being proposed under a much different set of conditions than existed when the 1998 scientific review paper recommendations were published. Prior to 1998, there was no official set of recommendations from a scientific body on eyewitness identification nor were there any guidelines from the legal system that were grounded in science.

Since the 1998 scientific review paper, however, numerous agencies, governing bodies, and organizations have proposed guidelines for collecting and preserving eyewitness identification evidence. These include the U.S. Department of Justice Guide (Technical Working Group on Eyewitness Evidence, 1999; reinforced recently by a directive from the United States Department of Justice, 2017), which followed closely on the 1998 scientific review paper as discussed in the introduction to the current article. Furthermore, a large set of U.S. state-wide reform guidelines have been enacted that range from the recommendations of state justice departments (e.g., Wisconsin), to executive orders from their Attorneys General (e.g., New Jersey), to court mandates (e.g., Oregon), to laws passed by state legislatures (e.g., North Carolina). In fact, in the United States there are now 21 states that require reforms through one or more of these mechanisms, namely California, Colorado, Connecticut, Florida, Georgia, Kansas, Illinois, Louisiana, Maryland, New Hampshire, Nebraska, Nevada, New Jersey, New York, North Carolina, Ohio, Oregon, Texas, Vermont, West Virginia, and Wisconsin. In other states, such as Montana, Hawaii, Minnesota, Pennsylvania, Iowa, Maine, Michigan, Washington, Idaho, and Rhode Island, police departments in major metropolitan areas have revamped their procedures. In effect, the reform of eyewitness identification procedures has been a chain-reaction going back to the 1998 scientific review paper, the Department of Justice guide in 1999, and the first state to adopt reforms (New Jersey in 2002). Each jurisdiction has modeled its reforms around those that preceded it and, hence, include the core four recommendations that were presented in the original 1998 scientific review paper (i.e., prelineup instructions, how to select fillers for the lineup, the need to collect a measure of witness confidence at the time of the identification, and the use of a double-blind procedure).

An important document relating to eyewitness identification procedures that was generated in recent years deserves special notice. The International Association of Chiefs of Police (IACP) produced a position article in 2016, in which they made specific recommendations regarding how lineups and showups should be conducted (International Association of Chiefs of Police, 2016). The IACP policy document goes beyond what most state reforms have done and goes beyond the recommendations in the 1998 scientific review paper. Specifically, the IACP recommendations not only include the four core recommendations, but also, like the current article, recommend avoiding multiple identifications procedures using the same witness and suspect, recommend video-recording the identification procedure, and make recommendations about showups that are similar to the recommendations that we offer in the current article. There are two recommendations in the current scientific review paper, however, that are not in the IACP policy, namely Recommendation 1 (Prelineup interview of the eyewitness) and Recommendation 2 (Evidence-based suspicion).

We note that the IACP policy document, like the policy documents that have been adopted in various states and localities, is almost exclusively a *prescriptive* document. Short, prescriptive documents can make sense for policy documents. The IACP policy, for example, is a mere four pages covering lineups and showups. Similarly, eyewitness identification policies and procedures documents adopted in various states and agencies are largely devoid of sustained rationale and explanations. In other words, these policy documents specify certain procedures, but they do not explain the rationale behind the procedures or review relevant science pertaining to the prescribed procedures. Hence, the current scientific review paper, like the 1998 scientific review paper, serves an important function that other policy documents do not, namely the function of documenting relevant science, where available, and an articulating rationale around each of the recommendations.

Earlier in this article we briefly alluded to an observation that deserves to be reemphasized. Although our nine recommendations are relatively specific and detailed, the most important aspects of our recommendations are the underlying principles associated with those recommendations. The double-blind administration of lineups, for example, is based on the principle that procedures must be in place to prevent the lineup administrator's knowledge of which person is the suspect to influence the eyewitness. Our specific preference for double-blind administration, however, is not the only way to achieve this prevention. With photo lineups, for instance, a properly programmed computer could administer the lineup without any administrator in the room at all. The underlying principle is important for another reason as well, namely that the specific recommendation might be technically followed and yet the principle itself is violated at some other point or level. Consider, for example, the invitation to view a lineup recommendation: "When inviting an eyewitness to attend a lineup procedure, police should not suggest that a suspect has been arrested or that the culprit will be present in the identification procedure." The underlying principle concerns the need to avoid communications to the eyewitness that could undermine the later prelineup instructions emphasizing that the culprit might not be in the lineup. Although the invitation to view a lineup might be the most likely time at which this principle is violated, it is not the only possible time. In this sense, the underlying principle is more important than the specific wording of the recommendation.

Finally, we think it is important for law enforcement to understand why certain procedures are recommended for constructing and conducting eyewitness identification procedures rather than simply being instructed on how to conduct those procedures. The current article provides valuable background on how research in areas such as perception, memory, decision making, and social influence can inform recommendations on lineup construction and presentation. Thus, in line with the National Research Council (2014) report on eyewitness evidence that highlights the importance of training for law enforcement, we believe that this article could serve as a useful resource for developing and implementing such training.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

References

Ackil, J. K., & Zaragoza, M. S. (1998). Memorial consequences of forced confabulation: Age differences in susceptibility to false memories. *Developmental Psychology, 34,* 1358–1372. http://dx.doi.org/10.1037/0012-1649.34.6.1358

Beaudry, J. L., Lindsay, R. C. L., Leach, A.-M., Mansour, J. K., Bertrand, M. I., & Kalmet, N. (2015). The effect of evidence type, identification accuracy, line-up presentation, and line-up administration on observers' perceptions of eyewitnesses. *Legal and Criminological Psychology, 20,* 343–364. http://dx.doi.org/10.1111/lcrp.12030

Behrman, B. W., & Davey, S. L. (2001). Eyewitness identification in actual criminal cases: An archival analysis. *Law and Human Behavior, 25,* 475–491. http://dx.doi.org/10.1023/A:1012840831846

Behrman, B. W., & Richards, R. E. (2005). Suspect/foil identification in actual crimes and in the laboratory: A reality monitoring analysis. *Law and Human Behavior, 29,* 279–301. http://dx.doi.org/10.1007/s10979-005-3617-y

Bergold, A. N., & Heaton, P. (2018). Does filler database size influence identification accuracy? *Law and Human Behavior, 42,* 227–243. http://dx.doi.org/10.1037/lhb0000289

Berkowitz, S. R., & Frenda, S. J. (2018). Rethinking the confident eyewitness: A reply to Wixted, Mickes, & Fisher. *Perspectives on Psychological Science, 13,* 336–338. http://dx.doi.org/10.1177/1745691617751883

Berkowitz, S. R., & Loftus, E. F. (2018). Misinformation in the courtroom. In H. Otgaar & M. L. Howe (Eds.), *Finding the truth in the courtroom: Dealing with deception, lies, and memories* (pp. 11–30). New York, NY: Oxford University Press.

Borchard, E. (1932). *Convicting the innocent: Errors of criminal justice.* New Haven, CT: Yale University Press.

Bradfield, A., & McQuiston, D. E. (2004). When does evidence of eyewitness confidence inflation affect judgments in a criminal trial? *Law and Human Behavior, 28,* 369–387. http://dx.doi.org/10.1023/B:LAHU.0000039331.54147.ff

Bradfield, A. L., Wells, G. L., & Olson, E. A. (2002). The damaging effect of confirming feedback on the relation between eyewitness certainty and identification accuracy. *Journal of Applied Psychology, 87,* 112–120. http://dx.doi.org/10.1037/0021-9010.87.1.112

Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory.* New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195154054.001.0001

Brandon, R., & Davies, C. (1973). *Wrongful imprisonment: Mistaken convictions and their consequences.* London, UK: George Allen and Unwin.

Brewer, N., Caon, A., Todd, C., & Weber, N. (2006). Eyewitness identification accuracy and response latency. *Law and Human Behavior, 30,* 31–50. http://dx.doi.org/10.1007/s10979-006-9002-7

Brewer, N., Keast, A., & Rishworth, A. (2002). The confidence-accuracy relationship in eyewitness identification: The effects of reflection and disconfirmation on correlation and calibration. *Journal of Experimental Psychology: Applied, 8,* 44–56. http://dx.doi.org/10.1037/1076-898X.8.1.44

Brewer, N., Keast, A., & Sauer, J. D. (2010). Children's eyewitness identification performance: Effects of a Not Sure response option and accuracy motivation. *Legal and Criminological Psychology, 15,* 261–277. http://dx.doi.org/10.1348/135532509X474822

Brewer, N., Vagadia, A. N., Hope, L., & Gabbert, F. (2018). Interviewing witnesses: Eliciting coarse-grain information. *Law and Human Behavior, 42,* 458–471. http://dx.doi.org/10.1037/lhb0000294

Brewer, N., Weber, N., & Semmler, C. (2005). Eyewitness identification. In N. Brewer & K. D. Williams (Eds.), *Psychology and law: An empirical perspective* (pp. 177–221). New York, NY: Guilford Press.

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied, 12,* 11–30. http://dx.doi.org/10.1037/1076-898X.12.1.11

Brown, C., Lloyd-Jones, T. J., & Robinson, M. (2008). Eliciting person descriptions from eyewitnesses: A survey of police perceptions of eyewitness performance and reported use of interview techniques. *The European Journal of Cognitive Psychology, 20,* 529–560. http://dx.doi.org/10.1080/09541440701728474

Brown, E., Deffenbacher, K., & Sturgill, W. (1977). Memory for faces and the circumstances of the encounter. *Journal of Applied Psychology, 62,* 311–318. http://dx.doi.org/10.1037/0021-9010.62.3.311

Carlson, C. A., Jones, A. R., Whittington, J. E., Lockamyeir, R. F., Carlson, M. A., & Wooten, A. R. (2019). Lineup fairness: Propitious heterogeneity and the diagnostic feature-detection hypothesis. *Cognitive Research: Principles and Implications, 4,* 20–26. http://dx.doi.org/10.1186/s41235-019-0172-5

Ceci, S. J., & Bruck, M. (1995). *Jeopardy in the courtroom: A scientific analysis of children's testimony.* Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/10180-000

Charman, S. D., Douglass, A. B., & Mook, A. (2019). Cognitive bias in forensic decision making. In N. Brewer & A. B. Douglass (Eds.), *Psychological science and the law* (pp. 30–53). New York, NY: Guilford Press.

Charman, S. D., Matuku, K., & Mook, A. (2019). Non-blind lineup administration biases administrator's interpretation of ambiguous witness statements and their perceptions of the witness. *Applied Cognitive Psychology, 33,* 1260–1270. http://dx.doi.org/10.1002/acp.3579

Charman, S. D., & Quiroz, V. (2016). Blind sequential lineup administration reduces both false identifications and confidence in those false identifications. *Law and Human Behavior, 40,* 477–487. http://dx.doi.org/10.1037/lhb0000197

Charman, S. D., & Wells, G. L. (2007). Eyewitness lineups: Is the appearance-change instruction a good idea? *Law and Human Behavior, 31,* 3–22. http://dx.doi.org/10.1007/s10979-006-9006-3

Charman, S. D., Wells, G. L., & Joy, S. W. (2011). The dud effect: Adding highly dissimilar fillers increases confidence in lineup identifications. *Law and Human Behavior, 35,* 479–500. http://dx.doi.org/10.1007/s10979-010-9261-1

Chase, W. G., & Simon, H. A. (1973). Perceptions in chess. *Cognitive Psychology, 4,* 55–81. http://dx.doi.org/10.1016/0010-0285(73)90004-2

Clark, S. E. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law and Human Behavior, 29,* 395–424. http://dx.doi.org/10.1007/s10979-005-5690-7

Clark, S. E. (2012a). Costs and benefits of eyewitness identification reform: Psychological science and public policy. *Perspectives on Psychological Science, 7,* 238–259. http://dx.doi.org/10.1177/1745691612439584

Clark, S. E. (2012b). Eyewitness identification reform: Data, theory, and due process. *Perspectives on Psychological Science, 7,* 279–283. http://dx.doi.org/10.1177/1745691612444136

Clark, S. E., Brower, G. L., Rosenthal, R., Hicks, J. M., & Moreland, M. B. (2013). Lineup administrator influences on eyewitness identification and eyewitness confidence. *Journal of Applied Research in Memory & Cognition, 2,* 158–165. http://dx.doi.org/10.1016/j.jarmac.2013.06.003

Clark, S. E., Howell, R. T., & Davey, S. L. (2008). Regularities in eyewitness identification. *Law and Human Behavior, 32,* 187–218. http://dx.doi.org/10.1007/s10979-006-9082-4

Clark, S. E., & Wells, G. L. (2008). On the diagnosticity of multiple-witness identifications. *Law and Human Behavior, 32,* 406–422. http://dx.doi.org/10.1007/s10979-007-9115-7

Clarke, C., Milne, R., & Bull, R. (2011). Interviewing suspects of crime: The impact of PEACE training, supervision and the presence of a legal advisor. *Journal of Investigative Psychology and Offender Profiling, 8,* 149–162. http://dx.doi.org/10.1002/jip.144

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

Clifford, B. R., & George, R. (1996). A field evaluation of training in three methods of witness/victim investigative interviewing. *Psychology, Crime & Law, 2,* 231–248. http://dx.doi.org/10.1080/10683169608409780

Colloff, M. F., Wade, K. A., & Strange, D. (2016). Unfair lineups make eyewitnesses more likely to confuse innocent and guilty suspects. *Psychological Science, 27,* 1227–1239. http://dx.doi.org/10.1177/0956797616655789

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.* Alexandria, VA: National Institute of Justice.

Corey, D., Malpass, R. S., & McQuiston, D. E. (1999). Parallelism in eyewitness and mock witness identifications. *Applied Cognitive Psychology, 13,* S41–S58.

Crombag, H. F., Wagenaar, W. A., & Van Koppen, P. J. (1996). Crashing memories and the problem of 'source monitoring'. *Applied Cognitive Psychology, 10,* 95–104.

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

Dando, C. J., Geiselman, R. E., MacLeod, N., & Griffiths, A. (2015). Interviewing adult witnesses and victims. In G. Oxburgh, T. Myklebust, T. Grant, & R. Milne (Eds.), *Communication in investigative and legal contexts: Integrated approaches from forensic psychology, linguistics and law enforcement* (pp. 79–106). Hoboken, NJ: Wiley-Blackwell. http://dx.doi.org/10.1002/9781118769133.ch5

Darling, S., Valentine, T., & Memon, A. (2008). Selection of lineup foils in operational contexts. *Applied Cognitive Psychology, 22,* 159–168. http://dx.doi.org/10.1002/acp.1366

Deffenbacher, K. A., Bornstein, B. H., McGorty, E. K., & Penrod, S. D. (2008). Forgetting the once-seen face: Estimating the strength of an eyewitness's memory representation. *Journal of Experimental Psychology: Applied, 14,* 139–150. http://dx.doi.org/10.1037/1076-898X.14.2.139

Deffenbacher, K. A., Bornstein, B. H., & Penrod, S. D. (2006). Mugshot exposure effects: Retroactive interference, source confusion, and unconscious transference. *Law and Human Behavior, 30,* 287–307. http://dx.doi.org/10.1007/s10979-006-9008-1

Dekle, D. J., Beal, C. R., Elliot, R., & Huneycutt, D. (1996). Children as witnesses: A comparison of lineup versus showup identification methods. *Applied Cognitive Psychology, 10,* 1–12. http://dx.doi.org/10.1002/(SICI)1099-0720(199602)10:1<1::AID-ACP354>3.0.CO;2-Y

Demarchi, S., & Py, J. (2009). A method to enhance person description: A field study. In R. Bull, T. Valentine, & T. Williamson (Eds.), *Handbook of psychology of investigative interviewing: Current developments and future directions* (pp. 241–256). Chichester, UK: Wiley. http://dx.doi.org/10.1002/9780470747599.ch14

Demarchi, S., Py, J., Groud-Than, S., Parain, T., & Brunel, M. (2013). Describing a face without overshadowing effect: Another benefice of the Person Description Interview. *Psychologie Française, 58,* 123–133. http://dx.doi.org/10.1016/j.psfr.2013.01.002

Doob, A. N., & Kirshenbaum, H. M. (1973). Bias in police lineups—Partial remembering. *Journal of Police Science and Administration, 1,* 287–293. http://dx.doi.org/10.1007/BF01039807

Douglass, A. B., Brewer, N., Semmler, C., Bustamante, L., & Hiley, A. (2013). The dynamic interaction between eyewitnesses and interviewers: The impact of differences in perspective on memory reports and interviewer behavior. *Law and Human Behavior, 37,* 290–301. http://dx.doi.org/10.1037/lhb0000034

Douglass, A. B., & Jones, E. E. (2013). Confidence inflation in eyewitnesses: Seeing is not believing. *Legal and Criminological Psychology, 18,* 152–167. http://dx.doi.org/10.1111/j.2044-8333.2011.02031.x

Douglass, A. B., Neuschatz, J. S., Imrich, J., & Wilkinson, M. (2010). Does post-identification feedback affect evaluations of eyewitness testimony and identification procedures? *Law and Human Behavior, 34,* 282–294. http://dx.doi.org/10.1007/s10979-009-9189-5

Douglass, A. B., & Smalarz, L. (2019). Post-identification feedback: State of the science. In M. Miller & B. Bornstein (Eds.), *Advances in psychology and law* (pp. 101–136). New York, NY: Springer. http://dx.doi.org/10.1007/978-3-030-11042-0_4

Douglass, A. B., Smith, C., & Fraser-Thill, R. (2005). A problem with double-blind photospread procedures: Photospread administrators use one eyewitness's confidence to influence the identification of another eyewitness. *Law and Human Behavior, 29,* 543–562. http://dx.doi.org/10.1007/s10979-005-6830-9

Dysart, J. E., Lawson, V. Z., & Rainey, A. (2012). Blind lineup administration as a prophylactic against the postidentification feedback effect. *Law and Human Behavior, 36,* 312–319. http://dx.doi.org/10.1037/h0093921

Dysart, J. E., Lindsay, R. C. L., Hammond, R., & Dupuis, P. (2001). Mug shot exposure prior to lineup identification: Interference, transference, and commitment effects. *Journal of Applied Psychology, 86,* 1280–1284. http://dx.doi.org/10.1037/0021-9010.86.6.1280

Eisen, M. L., Gabbert, F., Ying, R., & Williams, J. (2017). "I think he had a tattoo on his neck": How co-witness discussions about a perpetrator's description can affect eyewitness identification decisions. *Journal of Applied Research in Memory & Cognition, 6,* 274–282. http://dx.doi.org/10.1016/j.jarmac.2017.01.009

Eisen, M., Skerrit-Perta, A., Jones, J. M., Owen, J., & Cedre, G. C. (2017). Pre-admonition suggestion in live showups: When witnesses learn that the cops caught "the" guy. *Applied Cognitive Psychology, 31,* 520–529. http://dx.doi.org/10.1002/acp.3349

Eisen, M., Smith, A. M., Olaguez, A. P., & Skerritt-Perta, A. S. (2017). An examination of showups conducted by law enforcement using a field-simulation paradigm. *Psychology, Public Policy, and Law, 23,* 1–22. http://dx.doi.org/10.1037/law0000115

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1980). The deterioration of verbal descriptions of faces over different delay intervals. *Journal of Police Science and Administration, 8,* 101–106.

Fahsing, I. A., Ask, K., & Granhag, P. A. (2004). The man behind the mask: Accuracy and predictors of eyewitness offender descriptions. *Journal of Applied Psychology, 89,* 722–729. http://dx.doi.org/10.1037/0021-9010.89.4.722

Fawcett, J. M., Russell, E. J., Peace, K. A., & Christie, J. (2011). Of guns and geese: A meta-analytic review of the 'weapon focus' literature. *Psychology, Crime & Law, 19,* 35–66. http://dx.doi.org/10.1080/1068316X.2011.599325

Fisher, R. P., & Geiselman, R. E. (1992). *Memory-enhancing techniques for investigative interviewing: The cognitive interview.* Springfield, IL: Charles C Thomas.

Fisher, R. P., Schreiber Compo, N., Rivard, J., & Hirn, D. (2014). Interviewing witnesses. In T. J. Perfect & D. S. Lindsay's (Eds.), *The SAGE handbook of applied memory* (pp. 559–578). Los Angeles, CA: SAGE Publications. http://dx.doi.org/10.4135/9781446294703.n31

Fitzgerald, R. J., Price, H. L., Oriet, C., & Charman, S. D. (2013). The effect of suspect-filler similarity on eyewitness identification decisions: A meta-analysis. *Psychology, Public Policy, and Law, 19,* 151–164. http://dx.doi.org/10.1037/a0030618

Fitzgerald, R. J., Price, H. L., & Valentine, T. (2018). Eyewitness identification: Live, photo, and video lineups. *Psychology, Public Policy, and Law, 24,* 307–325. http://dx.doi.org/10.1037/law0000164

Flin, R. H., & Shepherd, J. W. (1986). Tall stories: Eyewitnesses' ability to estimate height and weight characteristics. *Human Learning: Journal of Practical Research & Applications, 5,* 29–38.

Flowe, H. D., Carline, A., & Karoglu, N. (2018). Testing the reflection assumption: A comparison of eyewitness ecology in the laboratory and criminal cases. *The International Journal of Evidence & Proof, 22,* 239–261. http://dx.doi.org/10.1177/1365712718782996

Flowe, H. D., Takarangi, M. K. T., Humphries, J. E., & Wright, D. S. (2016). Alcohol and remembering a hypothetical sexual assault: Can people who were under the influence of alcohol during the event provide accurate testimony? *Memory, 24,* 1042–1061. http://dx.doi.org/10.1080/09658211.2015.1064536

Frank, J., & Frank, B. (1957). *Not guilty.* London, UK: Gallanez.

Gabbert, F., & Brown, C. (2015). Interviewing for face identification. In T. Valentine & J. P. Davis (Eds.), *Forensic facial identification: Theory and practice of identification from eyewitnesses, composites and CCTV* (pp. 15–41). Chichester, UK: Wiley.

Gabbert, F., & Hope, L. (2013). Suggestibility and memory conformity. In A. M. Ridley, F. Gabbert, & D. J. La Rooy (Eds.), *Suggestibility in legal contexts: Psychological research and forensic implications* (pp. 63–84). London, UK: Wiley-Blackwell.

Gabbert, F., Hope, L., & Fisher, R. P. (2009). Protecting eyewitness evidence: Examining the efficacy of a self-administered interview tool. *Law and Human Behavior, 33,* 298–307. http://dx.doi.org/10.1007/s10979-008-9146-8

Gabbert, F., Memon, A., & Allan, K. (2003). Memory conformity: Can eyewitnesses influence each other's memories for an event? *Applied Cognitive Psychology, 17,* 533–543. http://dx.doi.org/10.1002/acp.885

Gabbert, F., Memon, A., Allan, K., & Wright, D. B. (2004). Say it to my face: Examining the effects of socially encountered misinformation. *Legal and Criminological Psychology, 9,* 215–227. http://dx.doi.org/10.1348/1355325041719428

Garrett, B. (2011). *Convicting the innocent: Where criminal prosecutions go wrong.* Cambridge, MA: Harvard University Press. http://dx.doi.org/10.4159/harvard.9780674060982

Garrioch, L., & Brimacombe, C. A. E. (2001). Lineup administrators' expectations: Their impact on eyewitness confidence. *Law and Human Behavior, 25,* 299–315. http://dx.doi.org/10.1023/A:1010750028643

Gobet, F., & Simon, H. A. (1996). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology, 31,* 1–40. http://dx.doi.org/10.1006/cogp.1996.0011

Godfrey, R. D., & Clark, S. E. (2010). Repeated eyewitness identification procedures: Memory, decision making, and probative value. *Law and Human Behavior, 34,* 241–258. http://dx.doi.org/10.1007/s10979-009-9187-7

Gonzalez, R., Ellsworth, P. C., & Pembroke, M. (1993). Misidentifications and failures to identify in lineups and showups. *Journal of Personality and Social Psychology, 64,* 525–537. http://dx.doi.org/10.1037/0022-3514.64.4.525

Goodsell, C. A., Neuschatz, J. S., & Gronlund, S. D. (2009). Effects of mugshot commitment on lineup performance in young and older adults. *Applied Cognitive Psychology, 23,* 788–803. http://dx.doi.org/10.1002/acp.1512

Granhag, P. A., Ask, K., & MacGiolla, E. (2013). Eyewitness recall: An overview of estimator-based research. In T. J. Perfect & D. S. Lindsay (Eds.), *SAGE handbook of applied memory* (pp. 35–55). Los Angeles, CA: SAGE Publications.

Granhag, P. A., Ask, K., Rebelius, A., Öhman, L., & MacGiolla, E. (2013). 'I saw the man who killed Anna Lindh!' An archival study of witnesses' offender descriptions. *Psychology, Crime & Law, 19,* 921–931. http://dx.doi.org/10.1080/1068316X.2012.719620

Greathouse, S. M., & Kovera, M. B. (2009). Instruction bias and lineup presentation moderate the effects of administrator knowledge on eyewitness identification. *Law and Human Behavior, 33,* 70–82. http://dx.doi.org/10.1007/s10979-008-9136-x

Greenberg, M. S., Westcott, D. R., & Bailey, S. E. (1998). When believing is seeing: The effect of scripts on eyewitness memory. *Law and Human Behavior, 22,* 685–694. http://dx.doi.org/10.1023/A:1025758807624

Gronlund, S. D., Carlson, C. A., Neuschatz, J. S., Goodsell, C. A., Wetmore, S., Wooten, A., & Graham, M. (2012). Showups versus lineups: An evaluation using ROC analysis. *Journal of Applied Research in Memory & Cognition, 1,* 221–228. http://dx.doi.org/10.1016/j.jarmac.2012.09.003

Gross, H. (1911). *Criminal psychology: A manual for judges, practitioners, and students.* Boston, MA: Little, Brown.

Harris, M. J., & Rosenthal, R. (1985). Mediation of interpersonal expectancy effects: 31 meta-analyses. *Psychological Bulletin, 97,* 363–386. http://dx.doi.org/10.1037/0033-2909.97.3.363

Haw, R. M., Dickinson, J. J., & Meissner, C. A. (2007). The phenomenology of carryover effects between show-up and line-up identification. *Memory, 15,* 117–127. http://dx.doi.org/10.1080/09658210601171672

Haw, R. M., & Fisher, R. P. (2004). Effects of administrator-witness contact on eyewitness identification accuracy. *Journal of Applied Psychology, 89,* 1106–1112. http://dx.doi.org/10.1037/0021-9010.89.6.1106

Hinz, T., & Pezdek, K. (2001). The effect of exposure to multiple lineups on face identification accuracy. *Law and Human Behavior, 25,* 185–198. http://dx.doi.org/10.1023/A:1005697431830

Hope, L., Gabbert, F., & Fisher, R. P. (2011). From laboratory to the street: Capturing witness memory using the Self-Administered Interview. *Legal and Criminological Psychology, 16,* 211–226. http://dx.doi.org/10.1111/j.2044-8333.2011.02015.x

Hope, L., Gabbert, F., Fisher, R. P., & Jamieson, K. (2014). Protecting and enhancing eyewitness memory: The impact of an initial recall attempt on performance in an investigative interview. *Applied Cognitive Psychology, 28,* 304–313. http://dx.doi.org/10.1002/acp.2984

Horry, R., Halford, P., Brewer, N., Milne, R., & Bull, R. (2014). Archival analyses of eyewitness identification test outcomes: What can they tell us about eyewitness memory? *Law and Human Behavior, 38,* 94–108. http://dx.doi.org/10.1037/h0093959

Horry, R., Memon, A., Wright, D. B., & Milne, R. (2012). Predictors of eyewitness identification decisions from video lineups in England: A field study. *Law and Human Behavior, 36,* 257–265. http://dx.doi.org/10.1037/h0093959

Huff, R., Rattner, A., & Sagarin, E. (1986). Guilty until proven innocent. *Crime and Delinquency, 32,* 518–544. http://dx.doi.org/10.1177/0011128786032004007

Innocence Project (2019). Retrieved from https://www.innocenceproject.org/all-cases/

International Association of Chiefs of Police. (2016). *Model policy on eyewitness identification.* Retrieved from https://www.theiacp.org/resources/policy-center-resource/eyewitness-identification

Jacoby, L. L., Woloshyn, V., & Kelley, C. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General, 118,* 115–125. http://dx.doi.org/10.1037/0096-3445.118.2.115

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3–28. http://dx.doi.org/10.1037/0033-2909.114.1.3

Jones, E. E., Williams, K. D., & Brewer, N. (2008). "I had a confidence epiphany!": Obstacles to combating post-identification confidence inflation. *Law and Human Behavior, 32,* 164–176. http://dx.doi.org/10.1007/s10979-007-9101-0

Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence-accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1304–1316. http://dx.doi.org/10.1037/0278-7393.22.5.1304

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237–251. http://dx.doi.org/10.1037/h0034747

Kaminski, K. S., & Sporer, S. L. (2017). Discriminating between correct and incorrect eyewitness identifications: The use of appropriate cues.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

Journal of Experimental Psychology: Applied, 23, 59–70. http://dx.doi.org/10.1037/xap0000110

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 336–358. http://dx.doi.org/10.1037/0278-7393.26.2.336

Kask, K., & Bull, R. (2009). From person descriptions to interviewing methods: What can be done to improve child witness testimonies? *Trames, 13,* 95–108. http://dx.doi.org/10.3176/tr.2009.2.01

Kassin, S. M. (1998). Eyewitness identification procedures: The fifth rule. *Law and Human Behavior, 22,* 649–653. http://dx.doi.org/10.1023/A:1025702722645

Kassin, S. M., Drizin, S. A., Grisso, T., Gudjonsson, G. H., Leo, R. A., & Redlich, A. D. (2010). Police-induced confessions: Risk factors and recommendations. *Law and Human Behavior, 34,* 3–38. http://dx.doi.org/10.1007/s10979-009-9188-6

Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory & Cognition, 2,* 42–52. http://dx.doi.org/10.1016/j.jarmac.2013.01.001

Kassin, S. M., Kukucka, J., Lawson, V. Z., & DeCarlo, J. (2014). Does video recording alter the behavior of police during interrogation? A mock crime-and-investigation study. *Law and Human Behavior, 38,* 73–83. http://dx.doi.org/10.1037/lhb0000047

Kassin, S. M., Kukucka, J., Lawson, V. Z., & DeCarlo, J. (2017). Police reports of mock suspect interrogations: A test of accuracy and perception. *Law and Human Behavior, 41,* 230–243. http://dx.doi.org/10.1037/lhb0000225

Kassin, S. M., Russano, M. B., Amrom, A. D., Hellgren, J., Kukucka, J., & Lawson, V. Z. (2019). Does video recording inhibit crime suspects? Evidence from a fully randomized field experiment. *Law and Human Behavior, 43,* 45–55. http://dx.doi.org/10.1037/lhb0000319

Keast, A., Brewer, N., & Wells, G. L. (2007). Children's metacognitive judgments in an eyewitness identification task. *Journal of Experimental Child Psychology, 97,* 286–314. http://dx.doi.org/10.1016/j.jecp.2007.01.007

Kebbell, M. R., & Milne, R. (1998). Police officers' perceptions of eyewitness performance in forensic investigations. *The Journal of Social Psychology, 138,* 323–330. http://dx.doi.org/10.1080/00224549809600384

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94,* 211–228. http://dx.doi.org/10.1037/0033-295X.94.2.211

Kleider, H. M., Pezdek, K., Goldinger, S. D., & Kirk, A. (2008). Schema-driven source misattribution errors: Remembering the expected from a witnessed event. *Applied Cognitive Psychology, 22,* 1–20. http://dx.doi.org/10.1002/acp.1361

Klobuchar, A., Steblay, N. M., & Caligiuri, H. L. (2006). Improving eyewitness identifications: Hennepin County's blind sequential lineup pilot project. *Cardozo Public Law, Policy, and Ethics Journal, 2,* 381–414.

Kocab, K., & Sporer, S. L. (2016). The weapon focus effect for person identifications and descriptions: A meta-analysis. In M. Miller & B. H. Bornstein (Eds.), *Advances in psychology and law* (Vol. 1, pp. 71–117). New York, NY: Springer. http://dx.doi.org/10.1007/978-3-319-29406-3_3

Kovera, M. B., & Evelo, A. J. (2017). The case for double-blind lineup administration. *Psychology, Public Policy, and Law, 23,* 421–437. http://dx.doi.org/10.1037/law0000139

Kovera, M. B., Penrod, S. D., Pappas, C., & Thill, D. L. (1997). Identification of computer-generated facial composites. *Journal of Applied Psychology, 82,* 235–246. http://dx.doi.org/10.1037/0021-9010.82.2.235

Kraus, U., Zeier, F., Wagner, W., Paelecke, M., & Hewig, J. S. (2017). Comparing the quality of memory reports in different initial eyewitness questioning approaches. *Cogent Psychology, 4,* 143–163. http://dx.doi.org/10.1080/23311908.2017.1403063

Lamb, M. E., Orbach, Y., Sternberg, K. J., Hershkowitz, I., & Horowitz, D. (2000). Accuracy of investigators' verbatim notes of their forensic interviews with alleged child abuse victims. *Law and Human Behavior, 24,* 699–708. http://dx.doi.org/10.1023/A:1005556404636

LaRooy, D., Brubacher, S. P., Aromäki-Stratos, A., Cyr, M., Hershkowitz, I., Korkman, J., & Stewart, H. (2015). The NICHD protocol: A review of an internationally-used evidence-based tool for training child forensic interviewers. *Journal of Criminological Research.* Policy & Practice, *1,* 76–89.

Lassiter, G. D. (2010). Psychological science and sound public policy: Video recording of custodial interrogations. *American Psychologist, 65,* 768–779. http://dx.doi.org/10.1037/0003-066X.65.8.768

Lassiter, G. D., Ware, L. J., Ratcliff, J. J., & Irvin, C. R. (2009). Evidence of the camera perspective bias in authentic videotaped interrogations: Implications for emerging reform in the criminal justice system. *Legal and Criminological Psychology, 14,* 157–170. http://dx.doi.org/10.1348/135532508X284293

Launay, C., & Py, J. (2015). Methods and aims of investigative interviewing of adult witnesses: An analysis of professional practices. *Pratiques Psychologiques, 21,* 55–70. http://dx.doi.org/10.1016/j.prps.2014.11.001

Lawson, V. Z., & Dysart, J. E. (2014). The showup identification procedure: An exploration of systematic biases. *Legal and Criminological Psychology, 19,* 54–68. http://dx.doi.org/10.1111/j.2044-8333.2012.02057.x

Levett, L. M. (2013). Co-witness information influences whether a witness is likely to choose from a lineup. *Legal and Criminological Psychology, 18,* 168–180. http://dx.doi.org/10.1111/j.2044-8333.2011.02033.x

Levi, A. M. (2011). Much better than the sequential lineup: A 120-person lineup. *Psychology, Crime & Law, 18,* 631–640. http://dx.doi.org/10.1080/1068316X.2010.526120

Levi, A. M., & Lindsay, R. C. L. (2001). Lineup and photo spread procedures: Issues concerning policy recommendations. *Psychology, Public Policy, and Law, 7,* 776–790. http://dx.doi.org/10.1037/1076-8971.7.4.776

Lindsay, R. C., Martin, R., & Webber, L. (1994). Default values in eyewitness descriptions. *Law and Human Behavior, 18,* 527–541. http://dx.doi.org/10.1007/BF01499172

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. http://dx.doi.org/10.1037/0021-9010.79.1.121

Lindsay, R. C. L., Wallbridge, H., & Drennan, D. (1987). Do clothes make the man? An exploration of the effect of lineup attire on eyewitness identification accuracy. *Canadian Journal of Behavioural Science, 19,* 463–478. http://dx.doi.org/10.1037/h0079998

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. http://dx.doi.org/10.1007/BF01040622

Loftus, E. F. (2004). Memories of things unseen. *Current Directions in Psychological Science, 13,* 145–147. http://dx.doi.org/10.1111/j.0963-7214.2004.00294.x

Loftus, E. F. (2017). Eavesdropping on memory. *Annual Review of Psychology, 68,* 1–18. http://dx.doi.org/10.1146/annurev-psych-010416-044138

Loftus, E. F., & Greene, E. (1980). Warning: Even memory for faces may be contagious. *Law and Human Behavior, 4,* 323–334. http://dx.doi.org/10.1007/BF01040624

Loftus, G. R., & Harley, E. M. (2005). Why is it easier to identify someone close than far away? *Psychonomic Bulletin & Review, 12,* 43–65. http://dx.doi.org/10.3758/BF03196348

Luus, C. A. E., & Wells, G. L. (1991). Eyewitness identification and the selection of distracters for lineups. *Law and Human Behavior, 15,* 43–57. http://dx.doi.org/10.1007/BF01044829

Malpass, R. S. (1981). Effective size and defendant bias in eyewitness identification lineups. *Law and Human Behavior, 5,* 299–309. http://dx.doi.org/10.1007/BF01044945

Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology, 66,* 482–489. http://dx.doi.org/10.1037/0021-9010.66.4.482

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87,* 252–271. http://dx.doi.org/10.1037/0033-295X.87.3.252

Manson v. Braithwaite, 432 U.S. 98 (1977).

Mansour, J. K., Beaudry, J. L., Kalmet, N., Bertrand, M. I., & Lindsay, R. C. L. (2017). Evaluating lineup fairness: Variations across methods and measures. *Law and Human Behavior, 41,* 103–115. http://dx.doi.org/10.1037/lhb0000203

Matuku, K., Douglass, A. B., & Charman, S. (2018, March). *A cautionary note about videotaped identification procedures.* Paper presented at the annual meeting of the American Psychology-Law Society, Memphis, TN.

McPhee, I., Paterson, H. M., & Kemp, R. I. (2014). The power of the spoken word: Can spoken recall enhance eyewitness evidence? *Psychiatry, Psychology and Law, 21,* 551–566. http://dx.doi.org/10.1080/13218719.2013.848001

Mecklenburg, S. H. (2006). *Report to the legislature of the State of Illinois: The Illinois pilot program on sequential double-blind identification procedures.* Retrieved from https://www.cga.ct.gov/jud/tfs/20130901_Eyewitness%20Identification%20Task%20Force/20110921/Illinois%20ID%20Pilot%20Program.pdf

Mecklenburg, S. H., Bailey, P. J., & Larson, M. R. (2008). The Illinois Field Study: A significant contribution to understanding real world eyewitness identification issues. *Law and Human Behavior, 32,* 22–27. http://dx.doi.org/10.1007/s10979-007-9108-6

Meissner, C. A. (2002). Applied aspects of the instructional bias effect in verbal overshadowing. *Applied Cognitive Psychology, 16,* 911–928. http://dx.doi.org/10.1002/acp.918

Meissner, C. A., Brigham, J. C., & Kelley, C. M. (2001). The influence of retrieval processes in verbal overshadowing. *Memory & Cognition, 29,* 176–186. http://dx.doi.org/10.3758/BF03195751

Meissner, C. A., Sporer, S. L., & Schooler, J. W. (2007). Person descriptions as eyewitness evidence. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of eyewitness psychology: Vol. II. Memory for people* (pp. 1–34). Mahwah, NJ: Erlbaum.

Memon, A., Gabbert, F., & Hope, L. (2004). The ageing eyewitness. In J. Adler (Ed.), *Forensic psychology: Debates, concepts and practice* (pp. 96–112). Devon, UK: Willan.

Memon, A., Havard, C., Clifford, B., Gabbert, F., & Watt, M. (2011). A field evaluation of the VIPER system: A new technique for eliciting eyewitness evidence. *Psychology, Crime & Law, 17,* 711–729. http://dx.doi.org/10.1080/10683160903524333

Memon, A., Meissner, C. A., & Fraser, J. (2010). The Cognitive Interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy, and Law, 16,* 340–372. http://dx.doi.org/10.1037/a0020518

Mickes, L. (2015). Receiver operating analysis and confidence-accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory & Cognition, 4,* 93–102. http://dx.doi.org/10.1016/j.jarmac.2015.01.003

Modjadidi, K., & Kovera, M. B. (2018). Viewing videotaped identification procedure increases juror sensitivity to single-blind photo-array administration. *Law and Human Behavior, 42,* 244–257. http://dx.doi.org/10.1037/lhb0000288

Molinaro, P. F., Arndorfer, A., & Charman, S. D. (2013). Appearance-change instruction effects on eyewitness lineup identification accuracy are not moderated by amount of appearance change. *Law and Human Behavior, 37,* 432–440. http://dx.doi.org/10.1037/lhb0000049

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology, 53,* 53–81. http://dx.doi.org/10.1146/annurev.psych.53.100901.135131

National Research Council. (2014). *Identifying the culprit: Assessing eyewitness identification.* Washington, DC: The National Academies Press. Retrieved from https://www.nap.edu/catalog/18891/identifyingthe-culprit-assessing-eyewitness-identification

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

Neuschatz, J. S., Wetmore, S. A., Key, K. N., Cash, D. K., Gronlund, S. D., & Goodsell, C. A. (2016). A comprehensive evaluation of showups. In B. Bornstein & M. K. Miller (Eds.), *Advances in psychology and law* (pp. 43–69). Cham, Switzerland: Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-29406-3_2

New Jersey v. Henderson, 27 A. 3d 872 (2011).

Newman, E. J., & Garry, M. (2013). False memory. In T. Perfect & D. S. Lindsay (Eds.), *SAGE handbook of applied memory* (pp. 110–126). Los Angeles, CA: SAGE Publications.

Oregon v. Lawson, 352 Or. 724 (2012).

Orfield, M. W., Jr. (1987). The exclusionary rule and deterrence: An empirical study of Chicago narcotics officers. *Chicano Law Review, 54,* 1016–1069. http://dx.doi.org/10.2307/1599834

Orfield, M. W., Jr. (1992). Deterrence, perjury, and the heater factor: An exclusionary rule in the Chicago criminal courts. *University of Colorado Law Review, 63,* 75–161.

Palmer, M. A., Brewer, N., & Weber, N. (2010). Postidentification feedback affects subsequent eyewitness identification performance. *Journal of Experimental Psychology: Applied, 16,* 387–398. http://dx.doi.org/10.1037/a0021034

Palmer, M. A., Brewer, N., Weber, N., & Nagesh, A. (2013). The confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention. *Journal of Experimental Psychology: Applied, 19,* 55–71. http://dx.doi.org/10.1037/a0031602

Perry v. New Hampshire, 565 U.S. 228 (2012).

Pezdek, K., & Blandon-Gitlin, I. (2005). When is an intervening lineup most likely to affect eyewitness identification accuracy? *Legal and Criminological Psychology, 10,* 247–263. http://dx.doi.org/10.1348/135532505X49846

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. http://dx.doi.org/10.1037/0021-9010.84.6.940

Pickel, K. L. (1998). Unusualness and threat as possible causes of "weapon focus". *Memory, 6,* 277–295. http://dx.doi.org/10.1080/741942361

Pickel, K. L. (1999). The influence of context on the "weapon focus" effect. *Law and Human Behavior, 23,* 299–311. http://dx.doi.org/10.1023/A:1022356431375

Police Executive Research Forum. (2013). *A national survey of eyewitness identification procedures in law enforcement agencies.* Retrieved from http://www.policeforum.org/assets/docs/Free_Online_Documents/Eyewitness_Identification/a national survey of eyewitness identification procedures in law enforcement agencies 2013.pdf

Pryke, S., Lindsay, R. C. L., Dysart, J. E., & Dupuis, P. (2004). Multiple independent identification decisions: A method of calibrating eyewitness identifications. *Journal of Applied Psychology, 89,* 73–84. http://dx.doi.org/10.1037/0021-9010.89.1.73

Quinlivan, D. S., Neuschatz, J. S., Cutler, B. L., Wells, G. L., McLung, J., & Harker, D. L. (2012). Do pre-admonition suggestions moderate the effect of unbiased lineup instructions? *Legal and Criminological Psychology, 17,* 165–176. http://dx.doi.org/10.1348/135532510X533554

Ratcliff, J. J., Lassiter, G. D., Schmidt, H. C., & Snyder, C. J. (2006). Camera perspective bias in videotaped confessions: Experimental evi-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

dence of its perceptual basis. *Journal of Experimental Psychology: Applied, 12,* 197–206. http://dx.doi.org/10.1037/1076-898X.12.4.197

Read, J. D., Yuille, J. C., & Tollestrup, P. (1992). Recollections of a robbery. *Law and Human Behavior, 16,* 425–446. http://dx.doi.org/10.1007/BF02352268

Reardon, M. C., & Fisher, R. P. (2011). Effect of viewing the interview and identification process on juror perceptions of eyewitness accuracy. *Applied Cognitive Psychology, 25,* 68–77. http://dx.doi.org/10.1002/acp.1643

Rhead, L. M., Rodriguez, D. N., Korobeynikov, V., Yip, J. H., & Kovera, M. B. (2015). The effects of lineup administrator influence and mortality salience on witness identification accuracy. *Journal of Forensic Psychology Practice, 15,* 248–274. http://dx.doi.org/10.1080/15228932.2015.1041362

Rodriguez, D. N., & Berry, M. A. (2014). The effect of line-up administrator blindness on the recording of eyewitness identification decisions. *Legal and Criminological Psychology, 19,* 69–79. http://dx.doi.org/10.1111/j.2044-8333.2012.02058.x

Rodriguez, D. N., & Berry, M. A. (2019). Administrator blindness affects the recording of eyewitness lineup outcomes. *Law and Human Behavior, 44,* 71–87. http://dx.doi.org/10.1037/lhb0000352

Rosenthal, R. (2002). Covert communication in classrooms, clinics, courtrooms, and cubicles. *American Psychologist, 57,* 839–849. http://dx.doi.org/10.1037/0003-066X.57.11.839

Ross, D. F., Ceci, S. J., Dunning, D., & Toglia, M. P. (1994). Unconscious transference and mistaken identity: When a witness misidentifies a familiar but innocent person. *Journal of Applied Psychology, 79,* 918–930. http://dx.doi.org/10.1037/0021-9010.79.6.918

Satin, G. E., & Fisher, R. P. (2019). Investigative utility of the Cognitive Interview: Describing and finding perpetrators. *Law and Human Behavior, 43,* 491–506. http://dx.doi.org/10.1037/lhb0000326

Sauer, J., Brewer, N., Zweck, T., & Weber, N. (2010). The effect of retention interval on the confidence-accuracy relationship for eyewitness identification. *Law and Human Behavior, 34,* 337–347. http://dx.doi.org/10.1007/s10979-009-9192-x

Sauerland, M., Krix, A. C., van Kan, N., Glunz, S., & Sak, A. (2014). Speaking is silver, writing is golden? The role of cognitive and social factors in written versus spoken witness accounts. *Memory & Cognition, 42,* 978–992. http://dx.doi.org/10.3758/s13421-014-0401-6

Sauerland, M., & Sporer, S. L. (2011). Written vs. spoken eyewitness accounts: Does modality of testing matter? *Behavioral Sciences & the Law, 29,* 846–857. http://dx.doi.org/10.1002/bsl.1013

Scheck, B., Neufeld, P., & Dwyer, J. (2000). *Actual innocence.* New York, NY: Random House.

Schreiber Compo, N., Carol, R. N., Evans, J. R., Pimentel, P., Holness, H., Nichols-Lopez, K., . . . Furton, K. G. (2017). Witness memory and alcohol: The effects of state-dependent recall. *Law and Human Behavior, 41,* 202–215. http://dx.doi.org/10.1037/lhb0000224

Shaw, J. S., III. (1996). Increases in eyewitness confidence resulting from post-event questioning. *Journal of Experimental Psychology: Applied, 2,* 126–146. http://dx.doi.org/10.1037/1076-898X.2.2.126

Shaw, J. S., III, Garven, S., & Wood, J. M. (1997). Co-witness information can have immediate effects on eyewitness memory reports. *Law and Human Behavior, 21,* 503–523. http://dx.doi.org/10.1023/A:1024875723399

Shaw, J. S., III, & McClure, K. A. (1996). Repeated post-event questioning can lead to elevated levels of eyewitness confidence. *Law and Human Behavior, 20,* 629–653. http://dx.doi.org/10.1007/BF01499235

Skov, R. B., & Sherman, S. J. (1986). Information-gathering processes: Diagnosticity, hypothesis-confirmatory strategies, and perceived hypothesis confirmation. *Journal of Experimental Social Psychology, 22,* 93–121. http://dx.doi.org/10.1016/0022-1031(86)90031-4

Slobogin, C. (1996). Testifying: Police perjury and what to do about it. *University of Colorado Law Review, 67,* 1037–1060.

Smalarz, L., Kornell, N., Vaughn, K. E., & Palmer, M. A. (2019). Identification performance from multiple lineups: Should eyewitnesses who pick fillers be burned? *Journal of Applied Research in Memory & Cognition, 8,* 221–232. http://dx.doi.org/10.1016/j.jarmac.2019.03.001

Smalarz, L., & Wells, G. L. (2014a). Confirming feedback following a mistaken identification impairs memory for the culprit. *Law and Human Behavior, 38,* 283–292. http://dx.doi.org/10.1037/lhb0000078

Smalarz, L., & Wells, G. L. (2014b). Post-identification feedback to eyewitnesses impairs evaluators' abilities to discriminate between accurate and mistaken testimony. *Law and Human Behavior, 38,* 194–202. http://dx.doi.org/10.1037/lhb0000067

Smith, A. M., Wells, G. L., Lindsay, R. C. L., & Myerson, T. (2018). Eyewitness identification performance on showups improves with an additional-opportunities instruction: Evidence for present-absent criteria discrepancy. *Law and Human Behavior, 42,* 215–226. http://dx.doi.org/10.1037/lhb0000284

Smith, A. M., Wells, G. L., Lindsay, R. C. L., & Penrod, S. D. (2017). Fair lineups are better than biased lineups and showups, but not because they increase underlying discriminability. *Law and Human Behavior, 41,* 127–145. http://dx.doi.org/10.1037/lhb0000219

Smith, A. M., Wells, G. L., Smalarz, L., & Lampinen, J. M. (2018). Increasing the similarity of lineup fillers to the suspect improves applied value of lineups without improving memory performance. *Psychological Science, 29,* 1548–1551. http://dx.doi.org/10.1177/0956797617698528

Smith, A. M., Wilford, M. M., Quigley-McBride, A., & Wells, G. L. (2019). Mistaken eyewitness identification rates increase when either witnessing or testing conditions get worse. *Law and Human Behavior, 43,* 358–368. http://dx.doi.org/10.1037/lhb0000334

Smith-Spark, J. H., Bartimus, J., & Wilcock, R. (2017). Mental time travel ability and the Mental Reinstatement of Context for crime witnesses. *Consciousness and Cognition, 48,* 1–10. http://dx.doi.org/10.1016/j.concog.2016.10.002

Snyder, M., & Swann, W. B. (1978). Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology, 36,* 1202–1212. http://dx.doi.org/10.1037/0022-3514.36.11.1202

Sporer, S. L. (1992). Post-dicting eyewitness accuracy: Confidence, decision-times and person descriptions of choosers and non-choosers. *European Journal of Social Psychology, 22,* 157–180. http://dx.doi.org/10.1002/ejsp.2420220205

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence and decision-times in simultaneous and sequential lineups. *Journal of Applied Psychology, 78,* 22–33. http://dx.doi.org/10.1037/0021-9010.78.1.22

Sporer, S. L. (1996). Psychological aspects of person descriptions. In S. L. Sporer, R. S. Malpass, & G. Koehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 53–86). Hillsdale, NJ: Erlbaum, Inc.

State v. Lawson, 352 Or. 724, 291 P. 3d 673 (2012).

Steblay, N. M. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. *Law and Human Behavior, 21,* 283–297. http://dx.doi.org/10.1023/A:1024890732059

Steblay, N. K. (2011). What we know now: The Evanston Illinois field lineups. *Law and Human Behavior, 35,* 1–12. http://dx.doi.org/10.1007/s10979-009-9207-7

Steblay, N. K. (2013). Lineup Instructions. In B. Cutler (Ed.), *Reform of eyewitness identification procedures* (pp. 65–86). Washington, DC: APA Press. http://dx.doi.org/10.1037/14094-004

Steblay, N. K., & Dysart, J. E. (2016). Repeated eyewitness identification procedures with the same suspect. *Journal of Applied Research in Memory & Cognition, 5,* 284–289. http://dx.doi.org/10.1016/j.jarmac.2016.06.010

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. http://dx.doi.org/10.1023/A:1025438223608

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

Steblay, N. K., Tix, R. W., & Benson, S. L. (2013). Double exposure: The effects of repeated identification lineups on eyewitness accuracy. *Applied Cognitive Psychology, 27,* 644–654. http://dx.doi.org/10.1002/acp.2944

Steblay, N., Wells, G. L., & Douglass, A. B. (2014). The eyewitness post identification feedback effect 15 years later: Theoretical and policy implications. *Psychology, Public Policy, and Law, 20,* 1–18. http://dx.doi.org/10.1037/law0000001

Sternberg, K. J., Lamb, M. E., Esplin, P. W., Orbach, Y., & Hershkowitz, I. (2002). Using a structure interview protocol to improve the quality of investigative interviews. In M. Eisen (Ed.), *Memory and suggestibility in the forensic interview* (pp. 409–436). Mahwah, NJ: Erlbaum.

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.

Tredoux, C. (1998). Statistical inference on measures of lineup fairness. *Law and Human Behavior, 22,* 217–237. http://dx.doi.org/10.1023/A:1025746220886

Tredoux, C., Parker, J. F., & Nunez, D. (2007). Predicting eyewitness identification accuracy with mock witness measures of lineup fairness: Quality of encoding interacts with line up format. *South African Journal of Psychology, 37,* 207–222. http://dx.doi.org/10.1177/008124630703700201

Tuckey, M. R., & Brewer, N. (2003). The influence of schemas, stimulus ambiguity, and interview schedule on eyewitness memory over time. *Journal of Experimental Psychology: Applied, 9,* 101–118. http://dx.doi.org/10.1037/1076-898X.9.2.101

Tunnicliff, J. L., & Clark, S. E. (2000). Selecting foils for identification lineups: Matching suspects or descriptions? *Law and Human Behavior, 24,* 231–258. http://dx.doi.org/10.1023/A:1005463020252

United States Department of Justice. (2017, January). *Eyewitness identification: Procedures for conducting photo arrays.* Memorandum for heads of department law enforcement components and all department prosecutors from Sally Yates, Deputy Attorney General, Washington, DC.

United States v. Ash, 413 U.S. 300 (1973).

United States vs. Wade, 388 U.S. 218, 234 (1967).

Valentine, T., Davis, J. P., Memon, A., & Roberts, A. (2012). Live show-ups and their influence on a subsequent video lineup. *Applied Cognitive Psychology, 26,* 1–23. http://dx.doi.org/10.1002/acp.1796

Valentine, T., Pickering, A., & Darling, S. (2003). Characteristics of eyewitness identification that predict the outcome of real lineups. *Applied Cognitive Psychology, 17,* 969–993. http://dx.doi.org/10.1002/acp.939

Vallano, J., Steele, L. E., Slapinski, K. A., Briggs, A. P., & Pozzulo, J. D. (2019). Familiar eyewitness identifications: The current state of affairs. *Psychology, Public Policy, and Law, 25,* 128–146. http://dx.doi.org/10.1037/law0000204

Van Koppen, P. J., & Lochun, S. K. (1997). Portraying perpetrators: The validity of offender descriptions by witnesses. *Law and Human Behavior, 21,* 661–685. http://dx.doi.org/10.1023/A:1024812831576

Vollmer, R. T. (2006). Predictive probability of serum prostate-specific antigen for prostate cancer: An approach using Bayes rule. *American Journal of Clinical Pathology, 125,* 336–342. http://dx.doi.org/10.1309/R5H6VUQ32KGJW448

Vredeveldt, A., Baddeley, A. D., & Hitch, G. J. (2012). The effects of eye-closure and "ear-closure" on recall of visual and auditory aspects of a criminal event. *Europe's Journal of Psychology, 8,* 284–299. http://dx.doi.org/10.5964/ejop.v8i2.472

Vredeveldt, A., Baddeley, A. D., & Hitch, G. J. (2014). The effectiveness of eye-closure in repeated interviews. *Legal and Criminological Psychology, 19,* 282–295. http://dx.doi.org/10.1111/lcrp.12013

Vredeveldt, A., Hildebrandt, A., & van Koppen, P. J. (2016). Acknowledge, repeat, rephrase, elaborate: Witnesses can help each other remember more. *Memory, 24,* 669–682. http://dx.doi.org/10.1080/09658211.2015.1042884

Wade, K. A., Nash, R. A., & Lindsay, D. S. (2018). Reasons to doubt the reliability of eyewitness memory: Commentary on Wixted, Mickes, and Fisher (2018). *Perspectives on Psychological Science, 13,* 339–342. http://dx.doi.org/10.1177/1745691618758261

Wagenaar, W. A., & van der Schrier, J. H. (1996). Face recognition as a function of distance and illumination: A practical tool for use in the courtroom. *Psychology, Crime & Law, 2,* 321–332. http://dx.doi.org/10.1080/10683169608409787

Walsh, D., & Bull, R. (2010). What really is effective in interviews with suspects? A study comparing interviewing skills against interviewing outcomes. *Legal and Criminological Psychology, 15,* 305–321. http://dx.doi.org/10.1348/135532509X463356

Weber, N., Brewer, N., Wells, G. L., Semmler, C., & Keast, A. (2004). Eyewitness identification accuracy and response latency: The unruly 10–12-second rule. *Journal of Experimental Psychology: Applied, 10,* 139–147. http://dx.doi.org/10.1037/1076-898X.10.3.139

Weber, N., & Perfect, T. J. (2012). Improving eyewitness identification accuracy by screening out those who say they don't know. *Law and Human Behavior, 36,* 28–36. http://dx.doi.org/10.1037/h0093976

Wells, G. L. (1978). Applied eyewitness testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology, 36,* 1546–1557. http://dx.doi.org/10.1037/0022-3514.36.12.1546

Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology, 14,* 89–103. http://dx.doi.org/10.1111/j.1559-1816.1984.tb02223.x

Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist, 48,* 553–571. http://dx.doi.org/10.1037/0003-066X.48.5.553

Wells, G. L. (2001). Eyewitness lineups: Data, theory, and policy. *Psychology, Public Policy, and Law, 7,* 791–801. http://dx.doi.org/10.1037/1076-8971.7.4.791

Wells, G. L. (2006). Eyewitness identification: Systemic reforms. *Wisconsin Law Review, 2006,* 615–643.

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. http://dx.doi.org/10.1037/0021-9010.83.3.360

Wells, G. L., & Bradfield, A. L. (1999). Measuring the goodness of lineups: Parameter estimation, question effects, and limits to the mock witness paradigm. *Applied Cognitive Psychology, 13,* 27–40.

Wells, G. L., Leippe, M. R., & Ostrom, T. M. (1979). Guidelines for empirically assessing the fairness of a lineup. *Law and Human Behavior, 3,* 285–293. http://dx.doi.org/10.1007/BF01039807

Wells, G. L., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin, 88,* 776–784. http://dx.doi.org/10.1037/0033-2909.88.3.776

Wells, G. L., & Luus, E. (1990). Police lineups as experiments: Social methodology as a framework for properly-conducted lineups. *Personality and Social Psychology Bulletin, 16,* 106–117. http://dx.doi.org/10.1177/0146167290161008

Wells, G. L., & Olson, E. A. (2002). Eyewitness identification: Information gain from incriminating and exonerating behaviors. *Journal of Experimental Psychology: Applied, 8,* 155–167. http://dx.doi.org/10.1037/1076-898X.8.3.155

Wells, G. L., & Quinlivan, D. S. (2009). Suggestive eyewitness identification procedures and the Supreme Court's reliability test in light of eyewitness science: 30 years later. *Law and Human Behavior, 33,* 1–24. http://dx.doi.org/10.1007/s10979-008-9130-3

Wells, G. L., Rydell, S. M., & Seelau, E. P. (1993). On the selection of distractors for eyewitness lineups. *Journal of Applied Psychology, 78,* 835–844. http://dx.doi.org/10.1037/0021-9010.78.5.835

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

Wells, G. L., Small, M., Penrod, S. J., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior, 22,* 603–647. http://dx.doi.org/10.1023/A:1025750605807

Wells, G. L., Smith, A. M., & Smalarz, L. (2015). ROC analysis of lineups obscures information that is critical for both theoretical understanding and applied purposes. *Journal of Applied Research in Memory & Cognition, 4,* 324–328. http://dx.doi.org/10.1016/j.jarmac.2015.08.010

Wells, G. L., Steblay, N. K., & Dysart, J. E. (2012). Eyewitness identification reforms: Are suggestiveness-induced hits and guesses true hits? *Perspectives on Psychological Science, 7,* 264–271. http://dx.doi.org/10.1177/1745691612443368

Wells, G. L., Steblay, N. K., & Dysart, J. E. (2015). Double-blind photo lineups using actual eyewitnesses: An experimental test of a sequential versus simultaneous lineup procedure. *Law and Human Behavior, 39,* 1–14. http://dx.doi.org/10.1037/lhb0000096

Wells, G. L., & Turtle, J. W. (1986). Eyewitness identification: The importance of lineup models. *Psychological Bulletin, 99,* 320–329. http://dx.doi.org/10.1037/0033-2909.99.3.320

Wells, G. L., Yang, Y., & Smalarz, L. (2015). Eyewitness identification: Bayesian information gain, base-rate effect equivalency curves, and reasonable suspicion. *Law and Human Behavior, 39,* 99–122. http://dx.doi.org/10.1037/lhb0000125

West, E., & Meterko, V. (2017). Innocence Project: DNA exonerations, 1989–2014: Review of data and findings from the first 25 years. *Albany Law Review, 79,* 717–795.

Wetmore, S. A., Neuschatz, J. S., Gronlund, S. D., Wooten, A., Goodsell, C. A., & Carlson, C. A. (2015). Effect of retention interval on showup and lineup performance. *Journal of Applied Research in Memory & Cognition, 4,* 8–14. http://dx.doi.org/10.1016/j.jarmac.2014.07.003

Wilford, M. M., & Wells, G. L. (2013). Eyewitness system variables: Revisiting the system-variable concept and the transfer of system variables to the legal system. In B. L. Cutler (Ed.), *Reform of eyewitness identification procedures* (pp. 23–43). Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/14094-002

Wise, R. A., Safer, M. A., & Maro, C. M. (2011). What U.S. law enforcement officers know and believe about eyewitness factors, eyewitness interviews, and identification procedures. *Applied Cognitive Psychology, 25,* 488–500. http://dx.doi.org/10.1002/acp.1717

Wixted, J. T., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review, 121,* 262–276. http://dx.doi.org/10.1037/a0035940

Wixted, J. T., Mickes, L., Dunn, J. C., Clark, S. E., & Wells, W. (2016). Estimating the reliability of eyewitness identifications from police lineups. *Proceedings of the National Academies of Science, 113,* 304–309. http://dx.doi.org/10.1073/pnas.1516814112

Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest, 18,* 10–65. http://dx.doi.org/10.1177/1529100616686966

Wogalter, M. S. (1991). Effects of post-exposure description and imaging on subsequent face recognition performance. *Proceedings of the Human Factors Society, 35,* 575–579. http://dx.doi.org/10.1518/107118191786754680

Wogalter, M. S. (1996). Describing faces from memory: Accuracy and effects on subsequent recognition performance. *Proceedings of the Human Factors Society, 40,* 536–540. http://dx.doi.org/10.1177/154193129604001103

Wright, D. B., & McDaid, A. T. (1996). Comparing system and estimator variables using data from real lineups. *Applied Cognitive Psychology, 10,* 75–84. http://dx.doi.org/10.1002/(SICI)1099-0720(199602)10:1<75::AID-ACP364>3.0.CO;2-E

Wright, D. B., Memon, A., Skagerberg, E. M., & Gabbert, F. (2009). When eyewitnesses talk. *Current Directions in Psychological Science, 18,* 174–178. http://dx.doi.org/10.1111/j.1467-8721.2009.01631.x

Wright, D. B., & Skagerberg, E. M. (2007). Postidentification feedback affects real eyewitnesses. *Psychological Science, 18,* 172–178. http://dx.doi.org/10.1111/j.1467-9280.2007.01868.x

Yarmey, A. D. (1986). Verbal, visual, and voice identification of a rape suspect under different levels of illumination. *Journal of Applied Psychology, 71,* 363–370. http://dx.doi.org/10.1037/0021-9010.71.3.363

Yarmey, A. D. (2000). Retrospective duration estimations for variant and invariant events in field situations. *Applied Cognitive Psychology, 14,* 45–57. http://dx.doi.org/10.1002/(SICI)1099-0720(200001)14:1<45::AID-ACP623>3.0.CO;2-U

Yarmey, A. D., Jacob, J., & Porter, A. (2002). Person recall in field settings. *Journal of Applied Social Psychology, 32,* 2354–2367. http://dx.doi.org/10.1111/j.1559-1816.2002.tb01866.x

Yarmey, A. D., Yarmey, M. J., & Yarmey, A. L. (1996). Accuracy of eyewitness identifications in showups and lineups. *Law and Human Behavior, 20,* 459–477. http://dx.doi.org/10.1007/BF01498981

Young v. Conway 715 F. 3d 79 (2d Cir. 2013).

Yuille, J. C., & Cutshall, J. L. (1986). A case study of eyewitness memory of a crime. *Journal of Applied Psychology, 71,* 291–301. http://dx.doi.org/10.1037/0021-9010.71.2.291

Yuille, J. C., & Tollestrup, P. A. (1990). Some effects of alcohol on eyewitness memory. *Journal of Applied Psychology, 75,* 268–273. http://dx.doi.org/10.1037/0021-9010.75.3.268

Zajac, R., & Henderson, N. (2009). Don't it make my brown eyes blue: Co-witness misinformation about a target's appearance can impair target-absent line-up performance. *Memory, 17,* 266–278. http://dx.doi.org/10.1080/09658210802623950

Zajac, R., & Karageorge, A. (2009). The wildcard: A simple technique for improving children's target-absent lineup performance. *Applied Cognitive Psychology, 23,* 358–368. http://dx.doi.org/10.1002/acp.1511

Zimmerman, D. M., Chorn, J. A., Rhead, L. M., Evelo, A. J., & Kovera, M. B. (2017). Memory strength and lineup presentation moderate effects of administrator influence on mistaken identifications. *Journal of Experimental Psychology: Applied, 23,* 460–473. http://dx.doi.org/10.1037/xap0000147

Received June 4, 2019
Revision received November 19, 2019
Accepted November 20, 2019 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
Content may be shared at no cost, but any requests to reuse this content in part or whole must go through the American Psychological Association.

EXHIBIT 4

Check for updates

Psychological Science in the Public Interest
2021, Vol. 22(1S) 1S–18S
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/15291006211026259
www.psychologicalscience.org/PSPI
SAGE

Test a Witness's Memory of a Suspect Only Once

John T. Wixted[1], Gary L. Wells[2], Elizabeth F. Loftus[3,4], and Brandon L. Garrett[5]
[1]Department of Psychology, University of California, San Diego; [2]Department of Psychology, Iowa State University; [3]Department of Psychological Science, University of California, Irvine; [4]Department of Criminology, Law and Society, University of California, Irvine; and [5]Duke University School of Law

Abstract
Eyewitness misidentifications are almost always made with high confidence in the courtroom. The courtroom is where eyewitnesses make their *last* identification of defendants suspected of (and charged with) committing a crime. But what did those same eyewitnesses do on the *first* identification test, conducted early in a police investigation? Despite testifying with high confidence in court, many eyewitnesses also testified that they had initially identified the suspect with low confidence or failed to identify the suspect at all. Presenting a lineup leaves the eyewitness with a memory trace of the faces in the lineup, including that of the suspect. As a result, the memory signal generated by the face of that suspect will be stronger on a later test involving the same witness, even if the suspect is innocent. In that sense, testing memory contaminates memory. These considerations underscore the importance of a newly proposed recommendation for conducting eyewitness identifications: *Avoid repeated identification procedures with the same witness and suspect.* This recommendation applies not only to additional tests conducted by police investigators but also to the final test conducted in the courtroom, in front of the judge and jury.

Keywords
eyewitness identification, wrongful convictions, malleability of memory

> No man ever steps in the same river twice, for it's not the same river and he's not the same man.
>
> —Heraclitus

In a court of law, a credible eyewitness who confidently identifies a defendant as the culprit of a crime is often thought to provide direct and powerful evidence of guilt. Indeed, judges have traditionally characterized a courtroom identification as having an "independent" and direct "source" in the witness's memory. Although underappreciated in the legal system, despite being almost universally understood by experimental psychologists, an eyewitness identification in court does *not* provide direct evidence of guilt. Nor is it independently sourced in the witness's memory. Instead, by the time of trial, an eyewitness's memory has almost invariably been contaminated by a variety of factors and is therefore highly error prone. As of today, 375 prisoners have been exonerated by DNA testing, 21 of whom were on death row, and it is now widely understood that eyewitness misidentifications contributed to about 70% of these wrongful convictions (Innocence Project, 2020).

Eyewitness misidentifications typically first occur during early stages of a police investigation (e.g., when a lineup is administered), long before trial. Decades of research have therefore focused on proper methods for conducting lineups so as to minimize initial misidentifications. As the relevant research has accumulated over the years, consensus science-based recommendations about proper eyewitness identification procedures have evolved accordingly. The set of guidelines set forth by

Corresponding Author:
John T. Wixted, Department of Psychology, University of California, San Diego
Email: jwixted@ucsd.edu

Wells et al. (2020) include a new recommendation that is the focus of this article. Specifically, Recommendation 8 is as follows: "Avoid repeated identification procedures with the same witness and suspect." In other words, test a witness's memory for a suspect only once.

Under the right conditions, the first eyewitness identification test can provide reliable information. According to a review of the literature, on an *initial* lineup identification test of *uncontaminated* memory conducted in accordance with current recommendations (i.e., when a pristine procedure is used), confidence can be a reliable indicator of accuracy (Wixted & Wells, 2017). That is, a high-confidence identification implies high accuracy, whereas a low-confidence identification implies low accuracy (the veritable definition of eyewitness reliability). How often pristine conditions prevail in the real world is unknown, but it is known that on the *first* test, eyewitness identification evidence is potentially reliable.

No later test provides more reliable information than the first test because memory is malleable (Davis & Loftus, 2018). That is, like other forms of forensic evidence, memory can be contaminated. Critically, the best chance to test uncontaminated memory is the first test because the very act of testing memory can contaminate it (Steblay & Dysart, 2016). The importance of testing a witness's memory for a suspect only once is hard to overemphasize because the failure to abide by that simple rule might account for a large proportion of the wrongful convictions overturned by DNA evidence (Garrett, 2011). Later, we detail specific research findings and representative real-world cases supporting these claims.

Implementing this newly proposed reform should be simple and straightforward because it involves no special training beyond educating police investigators, prosecutors, and judges about its compelling science-based rationale. The purpose of this article is to do just that. We begin by tracing the growing awareness of the importance of the recommendation to avoid repeated testing by briefly reviewing how consensus science-based guidelines for conducting eyewitness identification procedures have evolved over the years.

The Evolution of Guidelines Pertaining to Eyewitness Identification Procedures

Beginning in 1998, teams of scientists (sometimes working with law enforcement and legal practitioners) have been commissioned to draw up recommendations for conducting eyewitness identification procedures four times. The first guidelines were enumerated in a "white paper" (Wells et al., 1998) commissioned by the American Psychology-Law Society (APLS). That document provided four recommendations for conducting police lineups. A police lineup consists of one suspect and several physically similar fillers. Nowadays, the police typically use photo lineups instead of the live lineups that were once the norm (Police Executive Research Forum, 2013). The four recommendations in that white paper were as follows: (a) the lineup administrator should be blind to the identity of the suspect, (b) the eyewitnesses should be informed that the culprit may or may not be in the lineup, (c) the suspect should not stand out in the lineup (i.e., the lineup should be fair), and (d) a confidence statement should be obtained at the time an identification is made and before any feedback from the police. All of these recommendations remain in force today, but in 1998, the importance of testing memory only once was not yet apparent.

One year later, the National Institute of Justice (NIJ) issued another set of science-based guidelines (Technical Working Group for Eyewitness Evidence, 1999). Whereas the 1998 white paper focused on lineups, per se, the NIJ guidelines were much broader, providing recommendations for creating mug books and composites, for interviewing eyewitnesses, for conducting showups (which involve only the suspect), and for conducting lineups. The lineup recommendations were similar to those in the 1998 white paper, albeit with added specificity on some issues (e.g., the recommendation that at least five fillers be included in a lineup). Still, no mention was made about the special importance of testing a witness's memory for a suspect only once.

In 2013, a committee was appointed by the National Academy of Sciences to provide updated recommendations for eyewitness identification tests (National Research Council, 2014). Some of the new recommendations emphasized system-level issues such as training law-enforcement officers in eyewitness identification procedures and conducting pretrial judicial inquiries into the reliability of the eyewitness evidence. With regard to eyewitness identification procedures per se, they reiterated some of the earlier recommendations and added others, such as the recommendation that the eyewitness identification procedure be videotaped. Critically, they also added a new recommendation that reflected increased awareness of the importance of the initial identification. Specifically, their Recommendation 7 was as follows: "Make Juries Aware of Prior Identifications." In justifying this new recommendation, the committee wrote: "In-court confidence statements may also be less reliable than confidence judgments made at the time of an *initial* [emphasis added] out-of-court identification; as memory fails and/or confidence grows disproportionately" (p. 110). They also noted that "Eyewitness testimony is a type of evidence where (as with forms of forensic trace evidence) contamination may occur pre-trial" (p. 109). Contamination is the crux of the issue.

The next major development occurred when the APLS commissioned Wells et al. (2020) to update the 1998 white paper in light of what has been learned since that time. There are now nine recommendations, including such new recommendations as conducting an interview before the lineup (in part to warn the witnesses against attempting to identify the culprit on social media and elsewhere) and, as noted above, avoiding repeated identifications with the same witness and same suspect. The overarching reason to avoid repeated tests is that memory is malleable. The essential problem is that on a second test, an individual can look familiar because of the exposure during the first test, even when it is not the right person. Next, we consider how the field came to appreciate that fact and how it leads to the conclusion that law enforcement should avoid testing a witness's memory for a suspect more than one time.

Memory Is Malleable

Concerns about the malleability of memory can be traced back to at least Munsterberg (1908), but a scientific consensus about how easily memories can be modified—or even manufactured outright—did not begin to emerge until the mid-1970s. At that time, Loftus and Palmer (1974) and Loftus et al. (1978) reported the once surprising but now widely accepted finding that something as subtle as the nature of a question posed to an eyewitness can influence what the witness later remembers. Subsequent studies showed that people can even be induced to falsely remember entire events that never happened, such as being lost in a shopping mall as a child (Loftus & Pickrell, 1995) or that they were attacked by a vicious animal (Porter et al., 1999).

The examples summarized above pertain to memory tested by recall (i.e., recollecting details pertaining to a prior event), but eyewitness identification is a *recognition* memory test. As noted earlier, the malleability of memory has proven to be a particularly pernicious force on these tests, considering that many of the DNA exonerations involved eyewitnesses who incorrectly "recognized" the innocent suspect as the culprit. A striking example of memory contamination in the context of recognition memory was reported by Morgan et al. (2013). They conducted a study of military personnel who were confined to a mock prisoner-of-war camp during survival-school training. Each trainee experienced about 30 min of physically confrontational interrogation while alone in a room with an instructor. After the interrogation, the trainee was left alone in an isolation cell. Later, a member of the research team entered the cell and asked questions about the interrogator ("Did your interrogator give you anything to eat?") while showing the participant a photograph of a White male (the "foil"), thereby falsely implying that he was the interrogator.

Next, memory for the interrogator was tested using a nine-person, target-absent, simultaneous photo lineup. The photo lineup contained a picture of the foil but not the actual interrogator (i.e., it was a target-absent lineup in which the foil's face had been differentially familiarized under highly suggestive conditions). Participants who had not been exposed to the foil's face following interrogation identified the foil as the interrogator 15% of the time. By contrast, participants who had been exposed to the foil's face identified the foil as the interrogator a remarkable 84% of the time.

As the example presented above clearly indicates, testing memory with suggestive or otherwise improper procedures contaminates memory. A natural assumption might be that testing memory under optimal conditions (i.e., in accordance with current recommendations) would not have a contaminating effect. This may very well be true when the memory test in question consists of interviewing a witness about their recollection of details about the crime (a recall test) using a proper procedure such as the Cognitive Interview (Fisher & Geiselman, 1992). When questioned properly, witnesses tend to recall accurate information. Recalling accurate information does not contaminate memory. Indeed, it can reinforce it by making the retained information more durable than it otherwise would be, a memory-enhancing phenomenon known as "the testing effect" (e.g., Roediger & Karpicke, 2006).

Unfortunately, the same is not true when memory is tested using a recognition procedure such as a lineup. Even when using a pristine lineup procedure that happens to involve an innocent suspect, testing memory generally contaminates memory for that individual, thereby rendering any later recognition test prejudicial. Next, we consider some theoretical concepts derived from years of basic-science research to understand how and why that happens.

A Primer on the Theoretical Understanding of Recognition Memory

Several long-standing and influential theoretical considerations help to make sense of recognition memory: (a) encoding specificity, (b) similarity-based matching, (c) elaborative processing, (d) signal detection theory, and (e) the source-monitoring framework. These are

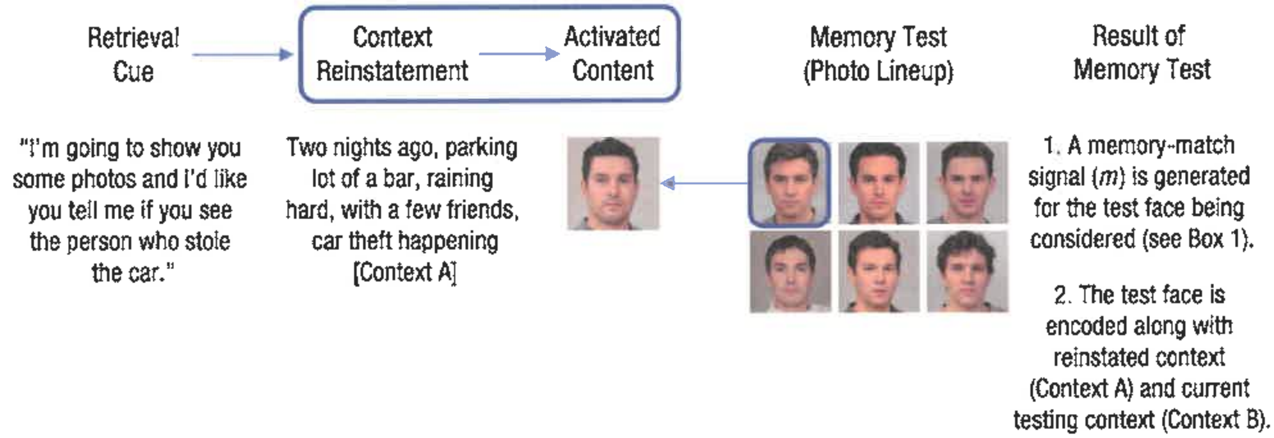


Fig. 1. The specific reference to the witnessed crime by the investigating officer (left column) is a retrieval cue that reinstates the encoding context in the mind of the eyewitness (Context A), which activates the relevant content—the face of the culprit. If multiple culprits were involved, their faces would also be activated, as would faces similar to those of the culprit(s) that might have been seen by the witness in other contexts. We omit those considerations for simplicity. Next, the lineup administrator presents the photo array to the witness, and each photo is compared with the activated content to make an identification decision. This figure illustrates that comparison process taking place for the top-left photo in the lineup. The comparison process yields a memory-match signal associated with the tested face (m) that is conceptualized in terms of signal detection theory (see Box 1). After all the faces in the array have been compared with memory, there will be six memory-match signals, and the face associated with the strongest signal will be a candidate for being identified. Comparing a face to memory involves elaborative processing and so incidentally creates a distinctive memory of the tested face (a process illustrated in Fig. 2), one that is encoded along with aspects of the reinstated context (Context A) and the testing context (Context B).

all standard "textbook" ideas that inform our understanding of the intuitively simple but surprisingly complex act of recognizing a once-seen face. The theoretical issues discussed in the remainder of this section are outlined in Figure 1.

Encoding specificity

Memory is generally understood to be cue-dependent (Tulving, 1983; Tulving & Thomson, 1973), which is to say that what you explicitly remember is determined by a *retrieval cue* that distinguishes the sought-after memory from the multitude of memories stored in one's brain. When memory is tested using a lineup, the retrieval cue consists of the specific question put to the witness. This is important because memories are differentially activated and thus accessible depending on the cues available at test (e.g., Godden & Baddeley, 1975). The question posed to the witness is not—or should not be—"are any of these faces familiar?" Instead, the more direct question is: "Do you see the person who committed the crime?" That retrieval cue will reinstate the context of the crime and activate the relevant content (i.e., the face of the culprit) in the brain of the eyewitness, as illustrated in the leftmost columns of Figure 1.

Similarity-based matching

In the simplest and perhaps most common case, the activated content consists of only one face (the singular culprit). If, instead, multiple culprits were involved, all their faces would be activated. According to global matching models, beginning with Gillund and Shiffrin (1984), each recognition-test item (e.g., each face in the lineup) is separately and individually compared against the activated faces (McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997). In the case of a single culprit, this process reduces to what one might already intuitively assume to be true: Each face in the lineup is separately compared against the remembered face of the culprit. Figure 1 illustrates the comparison process for one face in the lineup, which yields a memory-match signal (m) for that face. The more similar the face in the lineup is to the witness's memory of the culprit, the stronger this memory signal will be.

Signal detection theory

The memory-match signal (m) associated with a tested face is usually conceptualized in terms of signal detection theory. According to this theory, the memory signal is not all or none (i.e., match vs. no match) but is instead continuous because a face in the lineup can have any degree of similarity to the face of the culprit in memory. It seems natural to assume that, in a target-present lineup, the guilty suspect's face will be the most similar (generating the highest value of *m*), whereas the fillers will be less similar (generating weaker values of *m*). Likewise, in a fair target-absent lineup, no one will be very similar to the face of the culprit in memory, so they should all generate weak values of *m*. These assumptions are sensible, but they omit an important consideration. According to signal detection theory, only *on average* is a guilty suspect expected to generate a stronger memory signal than an innocent suspect or a filler. The reason is that memory matching is an inherently noisy process (Box 1). Thus, occasionally (but not usually), a guilty suspect will generate a weak memory signal and an innocent suspect (or a filler) will generate a strong memory signal. A troubling implication is that, even under ideal conditions involving no memory contamination and pristine testing procedures, and even on the initial test, misidentifications will inevitably happen from time to time. Still, high-confidence misidentifications should be rare. However, for reasons explained next, misidentifications would be expected to increase if memory is tested a second time.

Elaborative processing

The comparison process between a particular face in the lineup and the activated content of the culprit's face in memory does more than simply yield a memory-match signal. It also creates a detailed memory record because of the face processing that occurred during the identification procedure. In a typical lineup, the suspect and the fillers will be physically similar to each other. For example, to be included in the lineup, the face would ideally match the description of the perpetrator provided by the eyewitness (e.g., clean-shaven 20-year-old White male with short dark hair). Because of how lineups are designed, it will not suffice to perform a superficial scan of each face to make an identification decision, such as taking notice only of the shared features. Instead, each face in the lineup must be more thoroughly processed by attending to additional dimensions of the face (Fig. 2).

The act of attending to additional facial dimensions means that the witness has processed some of the unique features that, in configuration, define how a face in the lineup differs from other faces in the population. In other words, by necessity, a face in a lineup is *elaboratively processed* to decide whether this is the person who committed the crime. Such elaborative processing takes place whether the ultimate identification decision is "yes" or "no." Decades of research have established that the more elaboratively a stimulus is processed, the more likely it is to be later remembered (Craik & Tulving, 1975). Why? Craik (2002) put it this way: ". . . a richly elaborate trace will be more differentiated from other episodic records—this greater *distinctiveness* in turn will support more effective recollection in an analogous way to distinctive objects being more discriminable in the visual field" (pp. 306–307).

Elaborative processing creates a memory *incidentally* (i.e., without intention to form a memory). This is, in fact, the essence of the problem associated with testing a witness's memory for a suspect a second time. On that second test, a newly formed memory of the suspect will be accessible, even if the tested suspect is innocent, and the signal generated by the memory of the suspect's face might now be strong.

Source monitoring

The memory of a previously tested face is defined not only by its strength (i.e., by the magnitude of *m*) but also by the memory of the context that accompanied the encoding of the face. Assigning context to the memory signal is known as *source attribution*, and it can be an error-prone process (Johnson et al., 1993). That is, the witness might misattribute the strong memory signal to Source A when, in fact, the face was actually encountered in Source B. Testing memory for the first time using a police lineup almost seems tailor-made for inducing a source misattribution when memory is tested a second time.

An elaboratively processed face encoded during the initial test is not stored in a vacuum. Instead, it is encoded along with aspects of both the internal (i.e., reinstated) context and the external (i.e., testing) context (e.g., Cox & Shiffrin, 2017; Nelson & Shiffrin, 2013). These contexts are labeled Context A and Context B, respectively, in Figure 1. Critically, on the first test, only the culprit's face has been associated with the context of the crime (unlike any filler or any innocent suspect). However, on the second test, more faces will have been associated with that context, including the face of an innocent suspect being tested a second time. When Source A is again reinstated at the time of the second test ("Do you see the person who committed the crime?"), the activated content would now include not just the culprit's face but also the faces that were previously tested, including the face of the innocent suspect.

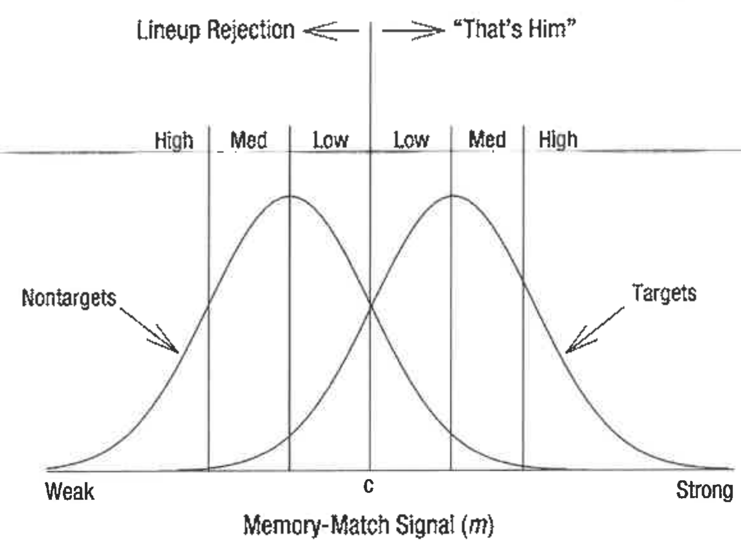
Typically, when the police conduct a second lineup with the same suspect and same witness, they use a

Box 1. Signal Detection Theory

Signal detection theory is a conceptual framework with origins dating back to the dawn of experimental psychology (Fechner, 1860/1966; Green & Swets, 1966; Kellen et al., 2021; Wixted, 2020). As applied to eyewitness identification, signal detection theory conceptualizes the memory-match signal (*m*) that is generated in the brain of an eyewitness when a face in a lineup (innocent or guilty) is compared with the face of the culprit stored in memory. The more similar the two faces are, the stronger the memory-match signal will be. On average, *m* will be strong when the face under consideration is the guilty suspect (the *target*) because that face matches the memory of the culprit, but it will not always be strong. This means that we should think of *m* not as a constant but as a variable that has a range of values across many eyewitnesses who are considering the guilty suspect in a target-present lineup. Its value will be high on average, but it will have variance as well. Thus, in signal detection theory, we represent *m* for guilty suspects as a *distribution* of values with a relatively high mean.

When the face under consideration is an innocent suspect in a target-absent lineup or a filler in either type of lineup (*nontargets*), *m* will be weak, on average, because these faces will not usually be particularly similar to the memory of the culprit. However, it will not always be weak (e.g., the innocent suspect might be a lookalike). Thus, once again, across many lineups and eyewitnesses, it is useful to conceptualize *m* as a variable with a relatively low mean, not as a constant. Thus, there are two distributions (assumed to be Gaussian in form and with equal variance for convenience) with different means, one for guilty suspects and another for both innocent suspects and fillers. For measurement purposes, we can conceptualize the difference between the target and nontarget means in standard deviation units and call that measure *d'* (a value estimated from data in a particular experiment).

After examining all the faces in the lineup, one face will have the maximum value of *m* (m_{max}). If m_{max} is strong enough—that is, if it exceeds the witness's *decision criterion* (c) for making an ID—that face is identified (Wixted et al., 2018). If so, the stronger m_{max} is, the higher the witness's confidence in that ID will be (low, medium or high). If even m_{max} is not strong enough to exceed the witness's decision criterion, the lineup is rejected. The weaker m_{max} is, the higher the witness's confidence will be that the culprit is not present in the lineup. This model inherently predicts a strong confidence–accuracy relationship (Wixted, 2020), which is often observed in lab studies for suspect IDs (Wixted & Wells, 2017) but is less reliably observed for lineup rejections (e.g., Brewer & Wells, 2006).



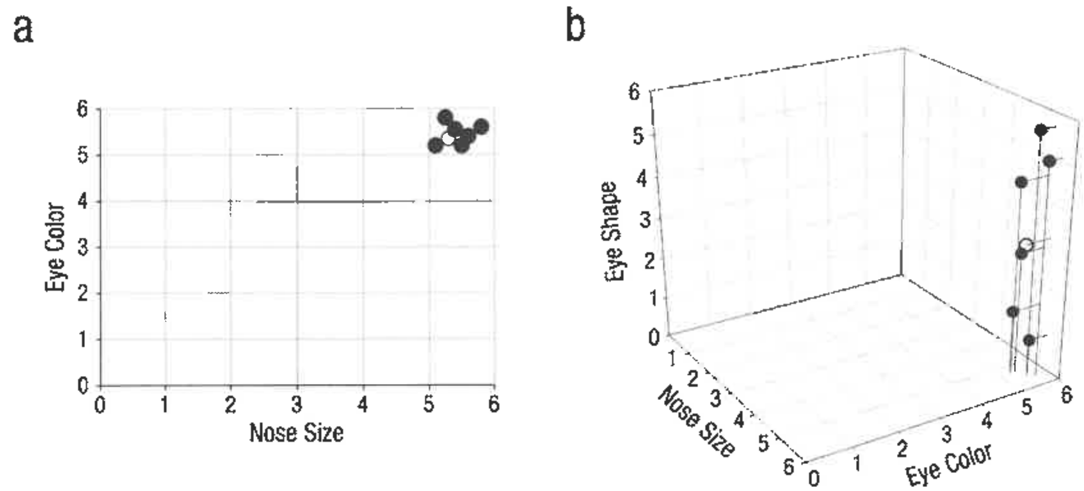


Fig. 2. Multidimensional "face space" (Valentine, 1991; Valentine et al., 2016). Although face perception and memory are widely thought to involve both holistic and feature-based processing (Abudarham et al., 2019; Chua et al., 2015; McKone & Yovel, 2009; Tanaka & Farah, 1993; Tanaka & Simonyi, 2016), low-level perceptual features are used as perceived facial dimensions here for illustrative purposes. The dimensions could just as easily reflect more global properties of any level of abstractness (e.g., masculinity, attractiveness, perceived trustworthiness). The graph in (a) shows perceptual representations of six members of a target-present lineup (filled circles) and the memory representation of the culprit (open circle) along two facial-feature dimensions, nose size and eye color. In this hypothetical example, all values fall between 5 and 6 on both dimensions because the witness described the culprit as having a large nose (0 = *very small* to 6 = *very large*) and dark brown eyes (0 = *very light blue* to 6 = *very dark brown*). The points cluster together because the lineup members were deliberately chosen to match this description of the culprit. When the points cluster together, as they would if only these two features were considered, it is hard for the witness to discriminate the guilty suspect from the fillers. Therefore, additional feature dimensions must be considered. The graph in (b) shows the perceptual and memory representations of the same individuals when a third feature dimension (eye shape) is considered (0 = *round* to 6 = *slanted*). The two-dimensional plot of eye color versus nose size in (a) is now the floor of the three-dimensional plot in (b); the points still cluster together on the floor. The vertical axis represents the new dimension (eye shape). Because this feature was not included in the witness's description, the lineup members exhibit natural variability, so the points spread out along this dimension. Moreover, because eye shape is a feature of the culprit's face that the witness stored in memory but did not describe, only the guilty suspect is now close to the memory representation of the culprit, which generates a differentially strong memory-match signal (m). The critical point here is that to make an identification decision, the witness has to consider additional feature dimensions beyond those included in the description. Critically, considering additional feature dimensions individuates a face and is an example of elaborative processing. Elaborative processing makes memories incidentally (i.e., without intention to form a memory).

new set of fillers. Therefore, in the typical case, the suspect will generate a *differentially* strong memory-match signal (potentially attributed to the wrong source) because only that face has been tested previously. The differential familiarization of the suspect's face when memory is tested a second time violates the basic tenets of the "lineups-as-experiments" analogy (Wells & Luus, 1990). The idea is that when police investigators conduct a lineup, they are essentially performing an experiment to test their hypothesis that the suspect is guilty. As experimenters, they should adopt the same protocols that scientists adopt to ensure the integrity of their experiments. One of those protocols is to ensure that participants (witnesses in this analogy) are blind to the hypothesized outcome lest they perform in such a way as to please the experimenter. But if the suspect is the only person in common between the first and the second identification tests, then it is clear to the witness which person the police suspect of having committed the crime—the person in common to both procedures. Because the witness is no longer blind to the suspect's identity on the second test, the lineup is inherently biased against the suspect.

Empirical Studies of Testing Memory a Second Time

In light of the foregoing theoretical considerations, the memory signal generated by the innocent suspect's face will likely be stronger on a second test involving the same witness as a result of the witness having observed the suspect on the first lineup test. The relevant empirical evidence unambiguously supports this theoretical prediction.

The memory-for-foils paradigm

An illuminating experimental design known as the *memory-for-foils* paradigm provides compelling evidence that testing memory contaminates memory by leaving behind a trace of the tested items (Jacoby et al., 2005). In a typical recognition memory experiment in the basic-science literature, participants are presented with a list of items to study (e.g., a list of words). On a later recognition memory test, those same items (now called "targets") are randomly intermixed with new items ("foils"), and each item is presented individually for a yes/no decision (i.e., "Did this item appear on the list, yes or no?"). Theoretically, the activated memory content against which each test item is compared consists of the items from the study list (Cox & Shiffrin, 2017). In a test like this, the targets are analogous to a guilty suspect because they were seen on the list, and the foils are analogous to innocent suspects and fillers because they were not seen on the list.

After completing the recognition test, the participants are then unexpectedly asked to complete a second recognition memory test consisting of the foils from the first test randomly intermixed with a new set of foils. This time, they are instructed to say "yes" to the foils that appeared on the first test (those items are now the targets) and to say "no" to the new foils. Theoretically, the activated memory set against which test items are compared consists of the items from the just-completed recognition memory test (including the foils). Therefore, the foils, when tested on the surprise memory test, will generate a relatively strong memory-match signal.

Indeed, participants perform very well on that second unexpected test even though, when they first saw the foils (now targets), they were merely attempting to decide whether or not those items had appeared on a previous list, not attempting to memorize them. The foils were elaboratively processed to answer the recognition memory question and were encoded incidentally.

This phenomenon is not limited to lists of words but occurs for faces tested in a lineup as well. In a study reported by Charman and Cahill (2012), participants first viewed a mock-crime video and were later tested using a standard six-person simultaneous photo lineup. Still later, the participants were given a surprise memory test for the five fillers in the lineup. This final test was a list memory test consisting of 10 faces (the five fillers plus five new faces), and each face was presented individually for a yes/no decision about whether it had been seen previously in the lineup. Keep in mind that during the lineup test, the participants were not attempting to memorize the faces; instead, they made only an identification decision about each face. On the final test, the results were striking: The hit rate (percentage of previously seen fillers correctly recognized as such) was 76%, whereas the false alarm rate (percentage of new faces incorrectly recognized as having been previously seen) was only 19%.[1]

It is worth briefly considering how these results are interpreted in terms of the standard signal detection model discussed earlier (Box 1) because it illustrates a key point about how memory contamination caused by testing memory should be conceptualized. The model holds that previously seen faces will generate a stronger memory signal, on average, than new faces. On the unexpected test, the memory signals generated by the new (previously unseen) fillers are conceptualized as having been drawn from a Gaussian distribution with a low mean (Fig. 3a). Because these memory signals are weak, on average, only a small percentage (19%) of them exceed the witness's decision criterion by chance. The memory signal generated by the previously seen fillers (i.e., fillers that became targets) are conceptualized as having been drawn from a Gaussian distribution with a high mean (Fig. 3b). Because these memory signals are strong, on average, a much higher percentage (76%) of them exceed the witness's decision criterion.

For these data, $d' = 1.58$ (Box 1), which means that the participants could easily discriminate previously seen fillers from new fillers. Of most relevance to the issue under consideration here is that this standard theoretical framework conceptualizes the memory distribution of the tested fillers (i.e., fillers that are now targets) as having been shifted upward relative to the fillers that had not yet been tested. In other words, it was not only the 76% of correctly recognized fillers that are now targets that had their memory signals strengthened by the initial test; the remaining 24% were strengthened (i.e., "contaminated") as well but not enough to exceed the decision criterion. Thus, the face was rejected on a second test, but perhaps with less confidence than would otherwise have been the case. The take-home message is that, in theory, testing memory contaminated memory for *all* the tested fillers.

Viewing mug shots

Brown et al. (1977) had participants observe two separate groups of five strangers ("criminals" hereafter). Ninety minutes later, they viewed 15 mug shots, including five people who were criminals and others who were being seen for the first time. A week later, the

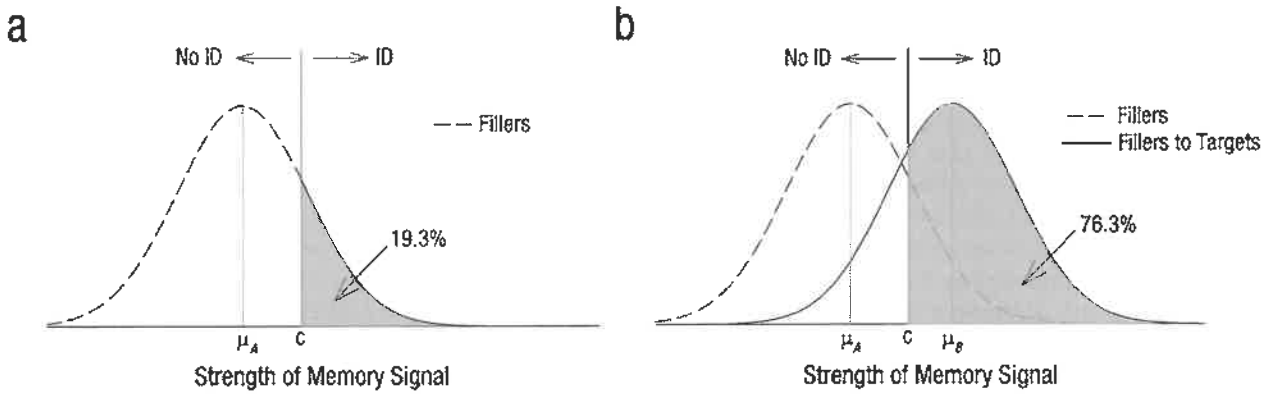


Fig. 3. Signal detection interpretation of findings reported by Charman and Cahill (2012). The graph in (a) represents the standard signal detection interpretation of a false alarm rate of 19.3%. The memory signals generated by new fillers have a mean of μ_A, and the decision criterion (c) is placed well above that. The graph in (b) shows the fillers-to-target distribution with mean μ_B superimposed on the information from (a), showing the standard signal detection interpretation of a hit rate of 76.3%. Note that, relative to new foils, all the fillers that are now targets (i.e., fillers to targets) have had their memory strengths boosted.

participants were presented with lineups (a second memory test) and asked to identify the criminals from the initial in-person encounter. The experience of viewing the mug shots had a clear effect on memory. For lineup members who had never been seen before, the rate of mistaken identification was 8%. However, if a lineup member's mug shot had been seen at the 90-min mark (but not during the original experience), the chances of being falsely identified as a criminal rose to 20%. Thus, the strong memory signals associated with the misidentified mug-shot-only faces were misattributed to the original experience involving the criminals (an example of source misattribution).

The findings reported by Brown et al. (1977) were reinforced by a study reported by Goodsell et al. (2015). Participants watched a short video clip of someone entering an office, after which they were randomly assigned to the mug-shot condition or to the no-mug-shot control condition. Those assigned to the mug-shot condition viewed 50 mug shots of people matched to the description of the culprit from the video, and they were asked to search for the perpetrator. All participants returned after a 48-hr delay and viewed either a target-present lineup or a target-absent lineup.

If a previously seen mug-shot photo in a lineup was one that the participant had previously picked, then that photo was (a) identified as the perpetrator from target-present lineups much more often than a photo of the actual perpetrator (.70 vs. .08) and (b) identified from target-absent lineups with a very high probability (.81). If the previously seen mug-shot photo was *not* the one that the participant had picked, then that photo was (a) *still* identified as the perpetrator from target-present lineups more often than a photo of the actual perpetrator (.28 vs. .18) and (b) identified from target-absent lineups with a probability of .38 (more than double the false-suspect ID rate from the control condition). Thus, memory was contaminated by the initial mug-shot test whether or not the mug-shot face appearing in the later lineup had been previously identified (see related findings reported by Memon et al., 2002).

Testing a suspect a second time

Conceptually similar effects are observed when the initial test consists of viewing a lineup rather than mug shots. In Steblay et al. (2013), participants viewed a video of a simulated crime and then attempted to identify the culprit from two six-person lineups separated by a 2-week retention interval. The suspect (guilty or innocent) was common to both lineups. In the absence of contamination from the first test, the expectation would be that the guilty-suspect ID rate would decline substantially after 2 weeks (because of forgetting) and the false ID rate would remain largely unchanged or increase slightly. For example, with similar retention intervals using a between-subjects design, Palmer et al. (2013) found the correct-ID rate dropped from .60 to .51 (immediate to delayed), $p = .052$, whereas a slight increase in the false-ID rate did not approach significance. By contrast, Steblay et al. (2013) found that when witnesses were tested both immediately and after a delay, the guilty-suspect ID rate *increased* on the delayed test, albeit nonsignificantly (instead of exhibiting the decrease expected as a result of forgetting), and the false-ID rate increased substantially from .21 to .31, $p = .03$. Thus, having seen the suspect in an earlier lineup contaminated memory, placing both innocent and guilty suspects at greater risk of being identified on a second test than would otherwise be the case.

Testing a witness's memory for a suspect a second time might not be problematic if, on the second test, not only was the same suspect included in the lineup but also the same fillers. In that case, everyone's face would generate an elevated memory signal compared with the first test, and no one would stand out. Lin et al. (2019) conducted this very experiment and found that, even then, nothing was gained by conducting the second test. Instead, witnesses simply became more willing to choose but without improving accuracy.

This Issue Is Specific to Forensic Memory Evidence

In the forensic context, the problem associated with repeated testing is specific to *memory* evidence. For example, repeatedly comparing latent fingerprints lifted from a crime scene to the known fingerprints of a suspect is not problematic and can even serve the cause of justice (e.g., fingerprint examiners can double-check their work) because the test itself does not change the evidence. By contrast, repeated tests of memory are unlikely to serve the cause of justice because testing changes memory (Wells et al., 2020). If comparing latent prints and known prints from a suspect altered the latent prints in such a way as to more closely resemble the fingerprints of the suspect, it seems reasonable to suppose that any fingerprint test after the first would be viewed with suspicion and perhaps excluded from consideration. Although this kind of contamination does not occur with fingerprints, it does occur with "face prints" (the memory of the culprit in the brain of the eyewitness).

Of course, as noted earlier, such contamination can occur even before the first official memory test conducted by the police, so it makes sense to take steps to prevent that from happening. In this regard, Recommendation 1 from Wells et al. (2020) is relevant. The recommendation is to conduct, before the lineup, an interview in which the witness is instructed to avoid attempting to identify the culprit on his or her own. If the witness has already done so and has encountered the suspect's photo (e.g., on social media), thereby contaminating memory before the first official test, it is also important to document that fact.

Memory Contamination Is Not the Only Problem

By focusing on memory contamination resulting from the initial test of memory, we do not mean to imply that it is the only problem associated with testing memory more than once—far from it. For example, as much prior research has shown, the risk to an innocent suspect associated with multiple testing is greatly compounded when suggestive procedures are used and/or when feedback to the witness is provided (Wells & Bradfield, 1998). If, for example, the witness misidentifies the innocent suspect with low confidence from a fair lineup, subsequent feedback from the police can quickly convert it to high confidence (e.g., if the police say "good job, we were pretty sure it was him"). In addition, other memory-contaminating events, such as seeing the face of the suspect again in pretrial hearings or in news stories will further strengthen the memory signal generated by the defendant's face at trial. In addition, if the witness discusses independent evidence against the suspect with prosecutors, it will help to cement the source misattribution according to which the strong memory signal reflects having originally seen the suspect commit the crime.

All of this would be avoided by testing memory only once, thereby strengthening the rationale for the new test-memory-once recommendation in Wells et al. (2020). The new point we are emphasizing here—one that has not received enough attention in the past—is that the witness's memory is ***already contaminated*** as a result of having taken the first test, even if pristine procedures were followed and even if none of the just-described additional factors exacerbated the problem (as difficult as that might be to imagine).

On the First Test, Confidence Protects Innocent Suspects

On the first (uncontaminated) test using a proper lineup, confidence is more likely to protect than imperil innocent suspects. As noted earlier (Box 1), signal detection theory predicts that decisions made with high confidence should be accurate most of the time, whereas decisions made with low confidence should more often be inaccurate. Related sequential sampling models (e.g., Pleskac & Busemeyer, 2010; Ratcliff, 1978; Ratcliff & Smith, 2004) make similar predictions about reaction time. That is, decisions made quickly should be accurate, whereas decisions made slowly should be less accurate. Empirically, these predictions have often been confirmed in list-memory studies conducted in the basic-science laboratory (e.g., Ratcliff & Murdock, 1976), in lineup studies conducted in the applied-science laboratory (Brewer et al., 2006), and in lineup studies conducted in the real world (Seale-Carlisle et al., 2019). For example, in a study involving actual eyewitnesses to a crime, Seale-Carlisle et al. (2019) reported that lineup decisions made rapidly (e.g., in 5 or 10 s) and with high confidence were estimated to be highly reliable, whereas decisions made slowly (e.g., 30 s or more) were much less reliable. This was true even in the rare case of a slow decision made with high confidence.

As unfortunate as a misidentification like this would be, keep in mind that the face of the innocent suspect does not actually correspond to the face stored in memory. Therefore, under optimal conditions, the strength of the memory-match signal, despite being randomly strong in a particular case, is not likely to far exceed the witness's decision criterion. According to standard assumptions of signal detection theory (Box 1), and in accordance with much empirical evidence, under ideal testing conditions, misidentifications of the innocent (and of fillers) are usually made with something other than high confidence (Fig. 4).

Losing sight of the low confidence that might be associated with an initial ID (and losing sight of other red flags, such as initial filler identifications or lineup rejections) wastes an opportunity to protect innocent suspects. Identifications made with low confidence are known to be highly error prone, which means that a low-confidence identification should be regarded as an inconclusive test outcome (Wixted & Wells, 2017). This is why the police record of an identification made with low confidence should never be written as "the witness positively identified the suspect." Instead, the record should reflect the lack of confidence, and that lack of confidence should be taken to mean that the memory test was inconclusive. The phrase "positively identified" is probably best reserved for cases where the witness is arguably *positive* that the identified individual is the culprit.

How Should Initial Confidence Be Measured?

The best way to determine whether the witness was "positive" is an actively researched issue, and there is no consensus. Fortunately, the available research suggests that the different methods (e.g., a verbal scale, a 5-point numerical scale, a 100-point numerical scale, asking the witness to use their own words) may not matter very much. For example, Tekin and Roediger (2017) tested 4-, 5-, 20-, and 100-point scales and found that the different scales yielded similar (continuous) confidence-accuracy plots. In their words, "the scales seem convertible from one to the other, and choice of scale range probably does not affect research into the relationship between confidence and accuracy" (p. 2).

Dodson and Dobolyi (2015) considered numerical compared with verbal confidence scales and concluded that "confidence is calibrated with accuracy in a nearly identical manner when confidence is expressed with either a numeric scale or a verbal scale" (p. 267). In agreement with this claim, Tekin et al. (2018) compared 2- and 4-point verbal and numeric scales and found little difference between them. Smalarz et al. (2021) and Mansour (2020) both asked participants to provide confidence in their own words or using a numerical scale. The results indicated that confidence was diagnostic of identification accuracy from a lineup either way, though Mansour (2020) found that verbal statements were more variable.

The upshot of the relevant research is that confidence should be assessed for an initial identification, as has been recommended for many years (Wells et al., 1998). Collecting a confidence statement of some kind appears to be more important than exactly how it is done. Critically, without a confidence statement, it is not possible to know whether the initial ID was made with low confidence, in which case it is highly error prone. Because the initial test is the one that matters, it is essential to collect a confidence statement on that first test. Moreover, the entire identification procedure should be recorded on video (Recommendation 7 of Wells et al., 2020) so that all interested parties—detective, prosecutor, judge, defense counsel, jury, and expert—can see and hear the confidence statement as it was captured in real time.

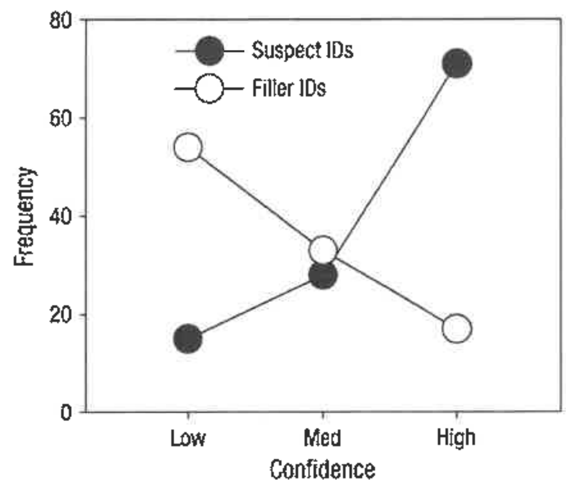


Fig. 4. Number of suspect IDs and filler IDs from 347 photo lineups administered to actual eyewitnesses in the Robbery Division of the Houston Police Department in 2013. The lineups were fair and were administered in double-blind fashion. Of interest here are the IDs made of known innocents (i.e., the fillers), the large majority of which were made with low or medium confidence. Reproduced from Fig. 1b of Wixted, J. T., Mickes, L., Dunn, J. C., Clark, S. E., & Wells, W. (2016). Estimating the reliability of eyewitness identifications from police lineups. *Proceedings of the National Academy of Sciences, USA, 113*, 304–309.

Plaintiff's Trial Exhibit 288
Page 11 of 18

What if the First Test Involves a Bad Photo of the Suspect?

When a witness fails to identify the suspect or does so with uncertainty on the first test, the police sometimes conclude that the photo of the suspect was not a very good likeness to the face of the suspect. Therefore, they try again, conducting a second test using what they believe to be a better photo. Does a second test endanger an innocent suspect under these conditions?

If the "bad" photo is far from perfect but is nonetheless recognizably the suspect, and if the witness elaboratively processed that bad photo to make a memory-based decision about it (perhaps choosing not to identify that individual), then the witness has processed features that individuate the suspect's face from other faces. The end product of such elaborative processing is an accessible memory record of that face. On the next memory test involving a better photo, those features will match the features that were encoded during the first test. This will have the effect of elevating the strength of the memory signal relative to what it otherwise would have been, thereby imperiling the innocent suspect. In other words, even a bad photo can contaminate memory if it is a recognizable photo of the suspect.

If the bad photo is instead not recognizable as the suspect, so much so that it might as well be a photo of a different person, then it is hard to see how memory contamination would occur. Therefore, in that case, a second test using a better photo would be reasonable. However, whether the photo is recognizable as the suspect is a judgment call. If no suspect ID is made, even a conscientious police investigator who strongly believes that the suspect is guilty might be inclined to honestly conclude that an imperfect-but-recognizable photo of the suspect was "bad," thereby justifying a second test. What can be done to protect against this alluring escape route from our main recommendation to test a witness's memory for a suspect only once?

The best solution would be to preserve the "bad" photo so that others can later judge for themselves whether it is a recognizable photo of the suspect. After all, this is not the only judgment call that an investigating officer has to make. The same officer will have judged the initial photo lineup to be fair, knowing that the photos would be preserved and later judged by others (e.g., by a jury at trial). Preserving the lineup photos incentivizes the investigating officer to exercise caution, ensuring that the lineup is fair. The same principle could be applied to the officer's judgment call about a photo of the suspect being so bad it might as well be a photo of another person. For example, at a pretrial hearing, if the court disagrees with that judgment call, then no later test involving the same suspect and eyewitness should be admissible as evidence.

In addition, the issue can be tested empirically. For example, a sample of people can be presented with the first (allegedly "bad") photo of the suspect and then asked if they can pick that person out from the second lineup. If they cannot do so with greater-than-chance accuracy, then it would be reasonable to conclude that the photo was in fact bad enough that it did not taint the second identification procedure. However, if people can pick the suspect out of the second lineup with greater-than-chance accuracy, then the second lineup should be suppressed.

How Important Was This Issue in the DNA-Exoneration Cases?

Wixted and Wells (2017) argued that on an initial test of *uncontaminated* memory using a *pristine* lineup procedure, high confidence can imply high accuracy, and low confidence can imply low accuracy. Opinions differ as to the reliability of high-confidence IDs in the real world (e.g., Do those tests typically involve uncontaminated memory? Do they typically involve pristine procedures?), but the consensus view is that low-confidence IDs are highly error prone (i.e., at best, they are only weakly probative of guilt). This is true whether or not memory has already been contaminated by the time of the initial test and whether or not pristine procedures are used. Moreover, filler IDs and lineup rejections on the first test are not neutral outcomes but, if anything, are probative of innocence (Wells & Lindsay, 1980).

With that background in mind, consider an analysis reported by Garrett (2011) in his book *Convicting the Innocent*. As noted earlier, data from the Innocence Project show that eyewitness misidentifications contributed to about 70% of more than 375 wrongful convictions later overturned by DNA evidence. Garrett (2011) analyzed the trial records from 161 of those cases in which an eyewitness misidentified an innocent suspect. In the courtroom, at trial, the eyewitness identifications were almost all made with high confidence, which makes sense (otherwise, the prosecutor likely would not have put the witness on the stand). What did these witnesses do at the time of the initial identification? We do not have contemporaneous records, but what these eyewitness and police witnesses described at trial, according to Garrett's (2011) analysis, yielded some interesting observations.

In 57% of the trial transcripts (92 of 161 cases), the witness who misidentified an innocent suspect with high confidence at trial recalled having initially done so with low confidence (34 cases); they recalled having identified a filler, another suspect, or no one at all (64 cases); or they reported not having seen the culprit's face (15 cases; some cases had more than one type of issue). We do not know what was said at these initial identifications, apart from what the witnesses later recounted at trial. However, to the extent that their recollections are accurate, these initial identifications were highly problematic, not only because the suspect was not confidently identified but for other reasons as well (many of these lineups also involved highly suggestive procedures). This is a problem because IDs made with low confidence are known to be highly error prone. As Garrett (2011) put it, this can provide "a glaring sign that the identification was not reliable" (p. 64). Low-confidence IDs, as well as nonidentifications, filler identifications, or identifications of other suspects, provide an opportunity to protect an innocent yet ultimately misidentified suspect. Unfortunately, for the DNA-exoneration cases involving an inconclusive outcome (or a contrary outcome) on the initial test, that opportunity was lost because the witness's memory was tested more than once.

How many of the remaining cases—the ones for which no testimony about the initial decision exists (43% of 161 cases)—also involved an initial outcome other than a high-confidence ID of the suspect? There is no way to know, in part because in only four cases was the procedure recorded; at a minimum, the evidence reported by Garrett (2011) is consistent with the idea that a sizable fraction of consequential eyewitness misidentifications began with something other than a conclusive (i.e., high-confidence) identification of the suspect. Indeed, as illustrated earlier in Figure 1, it is also consistent with the findings of a police department field study in which misidentifications of known innocents (fillers) were much more likely to have been made with low or medium confidence than high confidence. We turn now to three cases that illustrate how important this issue is.

Three Illustrative Cases

John Jerome White

On August 11, 1979, a man broke into a house in Manchester, Georgia, and raped a 74-year-old woman asleep on her couch (details about this case are available online: National Registry of Exonerations, 2016; "Understanding eyewitness misidentifications," 2011). On the basis of the description of the culprit provided by the victim, the police created a composite sketch. A Georgia Bureau of Investigation agent happened to be investigating a 19-year-old man named Jerome White on another charge, and he thought the sketch resembled White. A week later, the victim picked White out of a photo array, but she was not completely certain (saying she was "almost positive" he was the attacker).

Fig. 5. Live lineup administered to the victim following an initial photo lineup in which she identified Jerome White. White is in the middle, and the actual rapist (James Parham) is the man on the far right.

Perhaps because of those initial signs of uncertainty, she was later administered a live lineup (Fig. 5). White was the only person to appear in both the photo lineup and live lineup (none of the fillers were repeated), so his face had been differentially familiarized as a result of the initial photo-lineup test. Given all the theoretical and empirical considerations we have reviewed to this point, it would not be surprising to learn that the victim identified White again from this lineup, and she did. What makes this case remarkable, however, is who one of the fillers in the lineup turned out to be.

The victim originally told the police that her attacker was "well built" and had a "round face," a description that does not apply to Jerome White (Fig. 5). However, it does apply to one of the fillers in the lineup, namely, the man at the far right in Figure 5. He was not a suspect, but he happened to be in jail at the time, so he was selected to fill out the lineup. Incredibly, many years later, DNA evidence indicated that he was the one who actually committed the rape. Yet after seeing White's face in the initial photo lineup, choosing him, and (evidently) making a source misattribution, he was now the face that came to mind when the victim was asked if she saw the man who raped her in the lineup.

At his trial in 1980, the victim conclusively identified White as the man who had raped her ("that's him"). And why not? She had now seen his face on multiple previous tests, and the strong sense of familiarity was, in her mind, sourced to the initial crime (not to the lineup tests). The police and prosecutors presumably also reinforced her choice, not to intentionally create an injustice, but to reassure her. Unfortunately, such reassurance serves only to inflate confidence (Wells & Bradfield, 1998). Any doubts the witness had at the initial lineup vanished by the time of the trial. White spent more than 22 years in prison before finally being exonerated by DNA evidence in 2007. The same DNA evidence that exonerated White led prosecutors to charge James Parham (the man on the far right in Fig. 5) with the rape. He pleaded guilty and was sentenced to 20 years in prison.

Steven Gary Titus

On October 12, 1980, Port of Seattle police received a report that a man had raped a female hitchhiker—details about this case can be found from *The Seattle Times* (Henderson, 1981), and from a TED Talk (Loftus, 2013). Steve Titus was a restaurant manager in Seattle at the time, and he was on his way home from a restaurant with his fiancé when his car was stopped by a police officer because it resembled the car that was driven by the rapist. Titus also fit the description of the rapist provided by the victim. When later presented with a photo lineup, the victim identified Titus as her attacker, stating "This one is the closest one." It might very well be the case that Titus provided the closest match to her memory of the culprit (thereby generating the strongest memory signal of the faces in the lineup), but her wording is indicative of low confidence, not high confidence. Yet when Steve Titus was put on trial for rape, the witness's uncertainty had vanished. When she got on the witness stand, she identified Titus with high confidence. By then, not only did the witness have a memory of Titus based (at a minimum) on the initial lineup test, but she may also have been informed of other reasons why police and prosecutors thought he was guilty (further inflating confidence). Moreover, the courtroom identification test itself is inherently suggestive, inflating confidence still further. Based largely on that confident testimony, Titus was found guilty.

According to an article in *The New York Times* (Goleman, 1995), a few months after Titus was convicted, new evidence suggested that a different suspect was responsible for a series of rapes in the area. When the rape victim saw the photograph of the new suspect, she realized that he was the one who had actually raped her. At that point, she began to cry and said "Oh my God, what have I done to Mr. Titus?" However, the key mistake was made by other actors in the criminal-justice system, not the witness, because they tested memory for the suspect (Titus) more than once, ignoring her initial low-confidence ID. Instead of relying on the first test only, they unwittingly relied on contaminated memory evidence at trial to win what turned out to be a wrongful conviction.

Fortunately, Titus avoided a long stint in prison, but the story does not otherwise have a happy ending. Embittered by his wrongful conviction and the financial ruin it caused (including large legal fees and the loss of his job), he decided to file a lawsuit against the Port of Seattle police. Sadly, just before that case was to be heard, Titus died of heart failure at the age of 35.

Charles Don Flores

On the morning of January 29, 1998, witness Jill Barganier saw two people get out of a car outside the home of her neighbor, Elizabeth Black, who was murdered shortly thereafter. Details about this case can be found in legal documents posted at Flores's website (Flores, n.d.). Bargainer described both as White males with long, shoulder-length hair. Another neighbor independently described seeing two White males get out of the car and enter Black's house that morning. When presented with an initial photo lineup containing the main police suspect, a man named Richard Lynn Childs (a White man with long hair down to his shoulders), Barganier immediately identified him with high confidence. Childs owned a handgun of the same caliber used to murder Black, and he owned a conspicuously painted car that multiple eyewitnesses saw parked near Black's home the morning of the murder. He also eventually signed a confession admitting that he shot Black. This was an initial identification made quickly and with high confidence, and all indications are that it was a reliable ID.

Who was the other man Bargainer saw getting out the car that morning? The police suspected Charles Don Flores because he was a known associate of Childs and had been engaged in a drug deal with him in the hours before the murder. Flores was a heavyset Hispanic man with a crew cut and therefore did not match the description of the accomplice provided by the witness. Nevertheless, the police placed his photo in a lineup with other Hispanic men as fillers and presented it to the witness. Quite understandably, the witness did not identify anyone (i.e., she rejected the lineup). This makes sense because it is hard to see why photos of large Hispanic males with short hair would generate a strong memory-match signal compared against the memory of a White male with long hair stored in the witness's brain. Thus, on the initial test, her failure to identify Flores provides no evidence of guilt and instead provides evidence of innocence.

Nevertheless, at the trial, Jill Bargainer was certain that Flores was the man she saw that morning with Childs. Multiple factors presumably contributed to her high confidence, beginning with the memory-based decision she made about Flores on the first test (from that moment on, she likely had a representation of his face in her memory), perhaps continuing with news stories in which his face was shown, and culminating in the suggestive memory test performed at trial. For all these reasons, the only relevant eyewitness evidence was her rejection of the initial lineup.

In addition to the "direct" courtroom evidence provided by the eyewitness, there is indirect evidence against Flores as well. For example, in the days following the crime, he torched the conspicuous paint job on the car driven by Childs the morning of the murder (presumably to make it harder for the police to find), and he fled to Mexico when he learned that the police were looking for him (i.e., he "acted guilty"). This information, if Bargainer were aware of it, would have also served to bolster her confidence by the time of the trial. Childs did not testify about his accomplice at the time of the trial and has not done so to this day.[2]

Despite some independent evidence of guilt, by all accounts, it was the testimony of an extremely credible and highly confident eyewitness that led to the conviction of Flores. In Texas, murder is a capital crime, and an accomplice to a murder is as guilty as the triggerman (Childs). Therefore, Flores was sentenced to death. He has been on death row for over 21 years, and his appeal to the U.S. Supreme Court was denied on January 22, 2021. What may be his final appeal was recently filed in the Texas Court of Criminal Appeals.

The most remarkable fact about this case is that the eyewitness evidence that is mainly responsible for sending him to death row (namely, the witness's confident testimony at trial) is actually probative of innocence when properly understood (i.e., her initial description of the accomplice and her rejection of the initial lineup). In this case, police and prosecutors obviously failed to appreciate that only the first test counts.

A Simple Reform: Test a Witness's Memory for a Suspect Only Once

Presenting the face of a suspect on an eyewitness identification test contaminates the witness's memory for that individual. Such contamination is difficult to avoid, and if it occurs, there is no way to undo it (i.e., there is no way to decontaminate memory). If the witness's memory for that individual suspect is tested again, the suspect's face will generate a stronger memory signal than it otherwise would. As noted earlier, the fact that memory has been contaminated does not necessarily mean that the contaminated memory signal will be strong enough to exceed the criterion for making an identification. However, even in that case, memory has been irretrievably contaminated. Because of source misattribution, witnesses are at risk, on any later test, of responding to the elevated memory signal as if it were based on a memory formed at the time of the crime. By the time of trial, a variety of factors over and above the contaminating effects of testing memory more than once (feedback from the police, seeing the suspects face on the news, etc.) will have likely exacerbated the problem.

In contrast to this science-based perspective, judges often have a different legal perspective. As noted by Garrett (2012), they often embrace the catastrophically mistaken idea that, following the initial test, it is possible to conduct an "independent" test of memory, as if testing the match between a suspect's face and the witness's memory of the culprit multiple times is like testing the match between a suspect's fingerprints and the latent fingerprints lifted from a crime scene multiple times. However, as noted earlier, fingerprints do not change from the first test to the second; memories do. Therefore, once it has been tested and contaminated, it is not possible to perform a second independent test of the memory of a stranger's face that was formed during the commission of the crime.

The only barrier to implementing this recommended reform (testing a witness's memory for a suspect only once) is a faulty theory in the minds of various actors in the criminal-justice system. It therefore follows that implementing this reform should be much simpler than implementing other reforms that require training officers to administer eyewitness identification tests properly. To implement this newly proposed reform, the only training that is required is for policymakers to change their thinking about how memory works and to understand that testing memory for a suspect carries the high risk of irretrievably contaminating memory of that suspect. Considering how many wrongful convictions based on eyewitness misidentification might have been avoided by understanding this simple idea—and considering how many might be avoided going forward—implementing this reform should be an urgent priority.

A Final Word About Courtroom IDs

Because testing memory for a suspect is likely to contaminate memory for that face, a memory test conducted in the courtroom is likely to be a test of contaminated memory, by which time many additional factors exacerbate the problem. There may be rare exceptions (e.g., when the first test of memory for the defendant occurs from the witness stand, at trial, or when the prior test involved a photo that is not recognizable as the defendant). However, even in cases like that, despite avoiding the problem of memory contamination, a courtroom ID would still be problematic because of its inherently suggestive nature (Wells & Luus, 1990). It is inherently suggestive because only one person is sitting next to the defense attorney, making it plainly obvious to all that prosecutors believe they have enough evidence to be convinced that this is the person who committed the crime.

When it comes to eyewitness identifications, the courts often have it exactly backward, sometimes excluding earlier tests (including the all-important initial test) while allowing in court IDs based on memory that (unbeknownst to the judge) has likely been contaminated by events that occurred after the crime. As Garrett (2012) put it: "Today courts almost always allow courtroom identifications, but they sometimes bar prior identifications. Instead, courts should per se exclude courtroom identifications if there was a prior identification, but they should sometimes admit out-of-court identifications" (p. 457). Perhaps exceptions could be made for rare circumstances such as those mentioned above, but the point is that it makes sense for courts to exclude forensic evidence that has likely been contaminated in a way that is prejudicial to the defendant instead of making an exception for contaminated eyewitness evidence by routinely allowing it.

In addition to excluding courtroom identifications, except under presumably rare circumstances, the only out-of-court identification that should be admitted is the *first* one. Only the first test should be admitted for the same reason the court might exclude other kinds of forensic evidence that is likely to have been contaminated. Barring unusual circumstances (e.g., the witness did even look at the suspect on the first lineup test, or the photo used in the first lineup test was not even recognizable as the suspect), that first test provides the only relevant memory evidence. Even the first official test may involve contaminated memory (e.g., if the witness found a photo of the suspect on social media before viewing the photo lineup), but the first test unarguably provides the best chance to test uncontaminated memory. This simple reform, had it been implemented long ago, could have prevented many (perhaps most) of the wrongful convictions that occurred not because of eyewitness misidentification but because memory was tested more than once.

Transparency

Editor: Nora S. Newcombe

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Notes

1. The false alarm rate was not reported by Charman and Cahill (2012) because the focus of their analysis was different from ours, but the authors kindly provided us with the data.
2. In a plea bargain, Childs was sentenced to 30 years in prison and was released after serving 16 years (i.e., he is a free man today).

References

Abudarham, N., Shkillera, L., & Yovel, G. (2019). Critical features for face recognition. *Cognition, 182*, 73–83.

Brewer, N., Caon, A., Todd, C., & Weber, N. (2006). Eyewitness identification accuracy and response latency. *Law and Human Behavior, 30*, 31–50.

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied, 12*, 11–30.

Brown, E., Deffenbacher, K., & Sturgill, W. (1977). Memory for faces and the circumstances of encounter. *Journal of Applied Psychology, 62*, 311–318.

Charman, S. D., & Cahill, B. S. (2012). Witnesses' memories for lineup fillers postdicts their identification accuracy. *Journal of Applied Research in Memory and Cognition, 1*, 11–17.

Chua, K.-W., Richler, J. J., & Gauthier, I. (2015). Holistic processing from learned attention to parts. *Journal of Experimental Psychology: General, 144*, 723–729.

Cox, G. E., & Shiffrin, R. M. (2017). A dynamic approach to recognition memory. *Psychological Review, 124*, 795–860.

Craik, F. I. M. (2002). Levels of processing: Past, present . . . and future? *Memory, 10*(5-6), 305–318.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*, 268–294.

Davis, D., & Loftus, E. F. (2018). Eyewitness science in the 21st century: What do we know and where do we go from here? In J. Wixted (Ed.), *Stevens' handbook of experimental psychology and cognitive neuroscience: Vol. 1. Language and memory* (4th ed.). John Wiley & Sons. https://doi.org/10.1002/9781119170174.epcn116

Dodson, C. S., & Dobolyi, D. G. (2015). Misinterpreting eyewitness expressions of confidence: The featural justification effect. *Law and Human Behavior, 39*, 266–280.

Fechner, G. (1966). *Elements of psychophysics* (Vol. 1). Holt, Rinehart and Winston. (Original work published 1860)

Fisher, R. P., & Geiselman, R. E. (1992). *Memory-enhancing techniques for investigative interviewing: The cognitive interview.* Charles C Thomas.

Flores, C. D. (n.d.) *Legal documents.* https://charlesdflores.com/ressources/documents/

Garrett, B. L. (2011). *Convicting the innocent: Where criminal prosecutions go wrong.* Harvard University Press.

Garrett, B. L. (2012). Eyewitnesses and exclusion. *Vanderbilt Law Review, 65*, 451–506.

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

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology, 66*, 325–331.

Goleman, D. (1995, January 17). Studies point to flaws in lineups of suspects. *The New York Times*, C1. https://www.nytimes.com/1995/01/17/science/studies-point-to-flaws-in-lineups-of-suspects.html

Goodsell, C. A., Gronlund, S. D., & Neuschatz, J. S. (2015). Investigating mug shot commitment. *Psychology, Crime & Law, 21*, 219–233.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. John Wiley.

Understanding eyewitness misidentifications. (2011, March 14). Harvard University Press Blog. https://harvardpress.typepad.com/hup_publicity/2011/03/understanding-eyewitness-misidentifications.html

Henderson, P. (1981, May 15). One man's battle to clear his name. *The Seattle Times*. https://special.seattletimes.com/o/news/local/tituscase/index.html

Innocence Project. (2020). *The causes of wrongful conviction*. https://innocenceproject.org/causes-wrongful-conviction/.

Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychonomic Bulletin & Review, 12*, 852–857.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3–28.

Kellen, D., Winiger, S., Dunn, J. C., & Singmann, H. (2021). Testing the foundations of signal detection theory in recognition memory. *Psychological Review*. Advance online publication. https://doi.org/10.1037/rev0000288

Lin, W., Strube, M. J., & Roediger, H. L. (2019). The effects of repeated lineups and delay on eyewitness identification. *Cognitive Research: Principles and Implications, 4*, Article 16.

Loftus, E. (2013, March). How reliable is your memory? TED Conferences. https://www.ted.com/talks/elizabeth_loftus_how_reliable_is_your_memory?language=en

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 19–31.

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13*, 585–589.

Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals, 25*, 720–725.

Mansour, J. (2020). The confidence-accuracy relationship using scale versus other methods of assessing confidence. *Journal of Applied Research in Memory and Cognition, 9*, 215–231.

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.

McKone, E., & Yovel, G. (2009). Why does picture-plane inversion sometimes dissociate perception of features and spacing in faces, and sometimes not? Toward a new theory of holistic processing. *Psychonomic Bulletin Review, 16*, 778–797.

Memon, A., Hope, L., Bartlett, J., & Bull, R. (2002). Eyewitness recognition errors: The effects of mugshot viewing and choosing in young and old adults. *Memory & Cognition, 30*, 1219–1227.

Morgan, C. A., Southwick, S., Steffian, G., Hazlett, G. A., & Loftus, E. F. (2013). Misinformation can influence memory for recently experienced, highly stressful events. *International Journal of Law and Psychiatry, 36*, 11–17.

Munsterberg, H. (1908). *On the witness stand*. McClure.

National Registry of Exonerations. (2016, November 22). *John Jerome White*. https://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=3735

National Research Council. (2014). *Identifying the culprit: Assessing eyewitness identification*. The National Academies Press.

Nelson, A. B., & Shiffrin, R. M. (2013). The co-evolution of knowledge and event memory. *Psychological Review, 120*, 356–394.

Palmer, M. A., Brewer, N., Weber, N., & Nagesh, A. (2013). The confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention. *Journal of Experimental Psychology: Applied, 19*, 55–71.

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review, 117*, 864–901.

Police Executive Research Forum. (2013). *A national survey of eyewitness identification procedures in law enforcement agencies*. https://www.policeforum.org/assets/docs/Free_Online_Documents/Eyewitness_Identification/a%20national%20survey%20of%20eyewitness%20identification%20procedures%20in%20law%20enforcement%20agencies%202013.pdf

Porter, S., Yuille, J. C., & Lehman, D. R. (1999). The nature of real, implanted, and fabricated memories for emotional childhood events: Implications for the recovered memory debate. *Law and Human Behavior, 23*(5), 517–537.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*, 59–108.

Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review, 83*, 190–214.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review, 111*, 333–367.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long term-retention. *Psychological Science, 17*, 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Seale-Carlisle, T. M., Colloff, M. F., Flowe, H. D., Wells, W., Wixted, J. T., & Mickes, L. (2019). Confidence and response time as indicators of eyewitness identification accuracy in the lab and in the real world. *Journal of Applied Research in Memory and Cognition, 8*, 420–428.

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

Smalarz, L., Yang, Y., & Wells, G. L. (2021). Eyewitnesses' free-report verbal confidence statements are diagnostic of accuracy. *Law and Human Behavior*, *45*(2), 138-151.

Steblay, N. K., & Dysart, J. E. (2016). Repeated eyewitness identification procedures with the same suspect. *Journal of Applied Research in Memory and Cognition*, *5*, 284–289.

Steblay, N. K., Tix, R. W., & Benson, S. L. (2013). Double exposure: The effects of repeated identification lineups on eyewitness accuracy. *Applied Cognitive Psychology*, *27*, 644–654.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology*, *46A*, 225–245.

Tanaka, J. W., & Simonyi, D. (2016). The "parts and wholes" of face recognition: A review of the literature. *Quarterly Journal of Experimental Psychology*, *69*, 1876–1889.

Technical Working Group for Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement* (Report No. NCJ 178240). Offices of Justice Programs, U.S. Department of Justice. http://www.ncjrs.gov/pdffiles1/nij/178240.pdf

Tekin, E., Lin, W., & Roediger, H. L. (2018). The relationship between confidence and accuracy with verbal and verbal + numeric confidence scales. *Cognitive Research: Principles and Implications*, *3*, Article 41. https://doi.org/10.1186/s41235-018-0134-3

Tekin, E., & Roediger, H. L. (2017). The range of confidence scales does not affect the relationship between confidence and accuracy in recognition memory. *Cognitive Research: Principles and Implications*, *2*, Article 49.

Tulving, E. (1983). *Elements of episodic memory*. Clarendon Press.

Tulving, E., & Thomson, D. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352–373.

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

Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space: A unifying concept in face recognition research. *The Quarterly Journal of Experimental Psychology*, *69*, 1996–2019.

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., Kovera, M. B., Douglass, A. B., Brewer, N., Meissner, C. A., & Wixted, J. T. (2020). Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. *Law and Human Behavior*, *44*, 3–36.

Wells, G. L., & Lindsay, R. C. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin*, *88*(3), 776–784.

Wells, G. L., & Luus, C. E. (1990). Police lineups as experiments: Social methodology as a framework for properly conducted lineups. *Personality and Social Psychology Bulletin*, *16*, 106–117.

Wells, G. L., Small, M., Penrod, S. J., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior*, *22*, 603–647.

Wixted, J. T. (2020). The forgotten history of signal detection theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*, 201–233.

Wixted, J. T., Mickes, L., Dunn, J. C., Clark, S. E., & Wells, W. (2016). Estimating the reliability of eyewitness identifications from police lineups. *Proceedings of the National Academy of Sciences, USA*, *113*, 304–309.

Wixted, J. T., Vul, E., Mickes, L., & Wilson, B. W. (2018). Models of lineup memory. *Cognitive Psychology*, *105*, 81–114.

Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest*, *18*, 10–65. https://doi.org/10.1177/1529100616686966

EXHIBIT 5

Check for updates

Psychological Science in the Public Interest
2017, Vol. 18(1) 10–65
© The Author(s) 2017
Reprints and permissions: sagepub.com/journalsPermissions.nav
DOI: 10.1177/1529100616686966
www.psychologicalscience.org/PSPI
SAGE

The Relationship Between Eyewitness Confidence and Identification Accuracy: A New Synthesis

John T. Wixted[1] and Gary L. Wells[2]
[1]Department of Psychology, University of California, San Diego, and [2]Department of Psychology, Iowa State University

Summary
The U.S. legal system increasingly accepts the idea that the confidence expressed by an eyewitness who identified a suspect from a lineup provides little information as to the accuracy of that identification. There was a time when this pessimistic assessment was entirely reasonable because of the questionable eyewitness-identification procedures that police commonly employed. However, after more than 30 years of eyewitness-identification research, our understanding of how to properly conduct a lineup has evolved considerably, and the time seems ripe to ask how eyewitness confidence informs accuracy under more pristine testing conditions (e.g., initial, uncontaminated memory tests using fair lineups, with no lineup administrator influence, and with an immediate confidence statement). Under those conditions, mock-crime studies and police department field studies have consistently shown that, for adults, (a) confidence and accuracy are strongly related and (b) high-confidence suspect identifications are remarkably accurate. However, when certain non-pristine testing conditions prevail (e.g., when unfair lineups are used), the accuracy of even a high-confidence suspect ID is seriously compromised. Unfortunately, some jurisdictions have not yet made reforms that would create pristine testing conditions and, hence, our conclusions about the reliability of high-confidence identifications cannot yet be applied to those jurisdictions. However, understanding the information value of eyewitness confidence under pristine testing conditions can help the criminal justice system to simultaneously achieve both of its main objectives: to exonerate the innocent (by better appreciating that initial, low-confidence suspect identifications are error prone) and to convict the guilty (by better appreciating that initial, high-confidence suspect identifications are surprisingly accurate under proper testing conditions).

Keywords
calibration, confidence and accuracy, eyewitness identification, eyewitness memory, lineups, wrongful convictions

Introduction

In his book *On the Witness Stand: Essays on Psychology and Crime*, Hugo Münsterberg (1908) warned about the unreliability of eyewitness memory. As it turns out, he was prescient. Since 1989, 349 wrongful convictions have been overturned through DNA testing, and eyewitness misidentification played a role in over 70% of those cases—far more than any other contributing cause (Innocence Project, 2016). No one doubts that the large majority of these misidentifications were made in good faith. Somehow, these eyewitnesses came to honestly but mistakenly believe that the innocent defendant was the person who committed the crime. How did that happen? The short explanation is that the procedures used for testing eyewitness identification were not developed and validated in the scientific laboratory before being implemented in the field. Instead, they were developed

Corresponding Authors:
John T. Wixted, Department of Psychology, University of California, San Diego, La Jolla, CA 92093
E-mail: jwixted@ucsd.edu

Gary L. Wells, Psychology Department, West 112 Lagomarcino Hall, Iowa State University, Ames, IA 50021
E-mail: glwells@iastate.edu

within the criminal justice system and implemented under the mistaken assumption that they accurately identified the guilty without unduly jeopardizing the innocent.

When experimental psychologists began to empirically investigate the validity of these identification procedures in the 1970s and 1980s, they soon discovered that many seemed tailor-made for eliciting high-confidence misidentifications. For example, nowadays, a typical photo-lineup identification procedure consists of the simultaneous or sequential presentation of one photo of the suspect (the person the police believe may have committed the crime) and five or more fillers (photos of people who are known to be innocent but who physically resemble the suspect). Such a lineup offers protection to an innocent suspect because a witness who chooses randomly is far more likely to land on a filler than the suspect. However, before the dangers of eyewitness misidentification were understood, an investigating officer might present a lineup consisting of only suspects (with no fillers) and tell a witness who had just identified one of the suspects with low confidence that it was clearly the right decision, resulting in a higher expression of confidence the next time the witness was asked about it. By the time of the trial, the jury would see the witness honestly misidentify the suspect with high confidence and convict on that basis alone, often sending an innocent person to prison. Practices like these help to explain why, in every one of the DNA exoneration cases involving eyewitness misidentification examined by Garrett (2011), witnesses who mistakenly identified innocent defendants did so with high confidence when the case was tried in a court of law.

But what about the confidence expressed by an eyewitness tested using the scientifically validated procedures that have been developed over the years by eyewitness-identification researchers? That is the question we focus on here, and the answer will undoubtedly come as a surprise to many. Understandably, the disproportionate role played by eyewitness misidentification in the DNA exoneration cases has helped to create a widespread impression that eyewitness memory is unreliable even under the best of circumstances (i.e., that it is *inherently* unreliable). But over the last 20 years, eyewitness-identification researchers have discovered that when eyewitnesses are tested using appropriate identification procedures, the confidence they express can be, and usually is, a highly reliable indicator of accuracy (Brewer & Wells, 2006; Juslin, Olsson, & Winman, 1996). However, over that same period of time, the legal system has increasingly come to interpret the scientific literature as indicating no meaningful relation between confidence and accuracy. As a result, some courts now advise juries to disregard eyewitness expressions of confidence and to focus instead on a variety of other factors when trying to assess the reliability of an ID. The purpose of our article is to explain why a blanket disregard for eyewitness confidence not only is at odds with what has been learned in recent years but also can contribute both to the wrongful conviction of innocent suspects and to the unwarranted removal from suspicion of a guilty suspect.

Our article is organized as follows: We first document a growing trend within the legal system to disregard eyewitness confidence, with no distinction drawn as to whether the eyewitness-identification procedures were appropriate or not and with no distinction drawn between witness confidence at the time of the initial identification versus witness confidence at a later time. Next, we review a recommended set of appropriate ("pristine") identification procedures that have been developed in eyewitness-identification laboratory studies and how these pristine procedures can operate to prevent other factors from contaminating eyewitness confidence. The general idea is that a strong relation between confidence and accuracy is the natural state of affairs, but there are various things that can contaminate that relation. We then consider the nontrivial issue of how best to measure the confidence-accuracy relationship, followed by a detailed review and reanalysis of the empirical literature on the confidence-accuracy relation. The results will show that when pristine identification procedures are used, eyewitness confidence is a highly informative indicator of accuracy, and high-confidence suspect identifications are highly accurate. We go on to demonstrate that the confidence-accuracy relationship can be compromised when certain non-pristine identification procedures are used, and we enumerate priorities for future research on the confidence-accuracy relationship.

How Eyewitness Confidence Is Understood in the Legal System

In the legal system, eyewitness confidence is increasingly distrusted. For example, the state of New Jersey recently adopted jury instructions declaring that "although some research has found that highly confident witnesses are more likely to make accurate identifications, eyewitness confidence is generally an unreliable indicator of accuracy" (New Jersey Courts, 2012a; New Jersey Courts, 2012b). The report upon which the New Jersey instructions were based (Report of the Special Master, *State v. Henderson*, 2011) categorically asserted that "studies uniformly show, and the experts unanimously agree, that confidence is not closely correlated to accuracy" (p. 79). When discussing confidence, no distinction was drawn between identification procedures that are pristine and those that are not. These jury instructions are, of course, accurate when applied to problematic eyewitness-identification procedures, but our question concerns the confidence-accuracy relationship when pristine procedures are used early in the investigation and prior to any memory contamination. As our review will demonstrate, there are known conditions under which confidence clearly informs accuracy and other known conditions under which it clearly does not.

A bleak view of eyewitness confidence is not in any way limited to New Jersey. Other jurisdictions have revised their jury instructions so as to encourage juries to place little faith in eyewitness confidence. In Massachusetts, for example, the relevant instructions stipulate that "a witness's expressed certainty in an identification, standing alone, may not be a reliable indicator of the accuracy of the identification, especially where the witness did not describe that level of certainty when the witness first made the identification" (Massachusetts Court System, 2015, pp. 5–6). These instructions appropriately focus on the importance on the initial identification, but they do not appropriately communicate the high information value of an initial statement of confidence obtained from a pristine identification procedure.

Next, consider this recent statement made by the Connecticut Supreme Court in *State v. Guilbert* (2012): "Courts across the country now accept that there is at best a weak correlation between a witness's confidence in his or her identification and its accuracy." In a subsequent case, the Connecticut Psychological Association filed an amicus brief with the state supreme court arguing that eyewitness confidence is so loosely correlated with accuracy that it should no longer serve as a criterion for evaluating the reliability of eyewitness identification (Berard, 2014). No distinction was made between the confidence of the witness at the time of identification and the confidence of the witness at trial. Similarly, in *Brodes v. State* (2005), the Georgia Supreme Court held that jury instructions should not encourage jurors to consider a witness's confidence when trying to determine the reliability of an ID, specifically citing scientific research on the correlation between confidence and accuracy:

> In light of the scientifically-documented lack of correlation between a witness's certainty in his or her identification of someone as the perpetrator of a crime and the accuracy of that identification, and the critical importance of accurate jury instructions as "the lamp to guide the jury's feet in journeying through the testimony in search of a legal verdict," we can no longer endorse an instruction authorizing jurors to consider the witness's certainty in his/her identification as a factor to be used in deciding the reliability of that identification. Accordingly, we advise trial courts to refrain from informing jurors they may consider a witness's level of certainty when instructing them on the factors that may be considered in deciding the reliability of that identification.

Again, no distinction was made by the court between the confidence of the witness at the time of identification and the confidence of the witness at trial. Along the same lines, in *State v. Mitchell* (2012), the Utah Supreme Court recently stated,

> In the end, we agree with the Connecticut Supreme Court that the available studies are not definitive on the question whether there is a significant correlation between certainty and accuracy. But we are also mindful that the literature suggests certainty may not always be as reliable an indicator of accuracy. . . Therefore, we hold it is error to instruct the jury on the degree of certainty factor, and we discourage its future use.

Undeniably, eyewitness certainty at pretrial hearings or at trial should be highly suspect for reasons we will discuss. But when a lineup is conducted under pristine testing conditions and the confidence statement of the witness is taken at the time of identification, the data indicate that confidence is a reliable indicator of accuracy.

The fact that courts increasingly distrust eyewitness confidence is not altogether surprising, given that expert witnesses and concerned organizations routinely paint a gloomy picture of the confidence-accuracy relationship. For example, a 2013 amicus brief filed by the Innocence Project said, "A witness's confidence bears, at best, a weak relationship to accuracy" (Innocence Project, 2013, p. 11). However, the evidence we will review suggests that "at best" (i.e., under pristine conditions), a witness's confidence bears a strong relationship to accuracy.

It is not just the Innocence Project that has a generally pessimistic view of the confidence-accuracy relationship. A recent amicus brief filed by the American Psychological Association painted a similarly bleak picture of the situation:

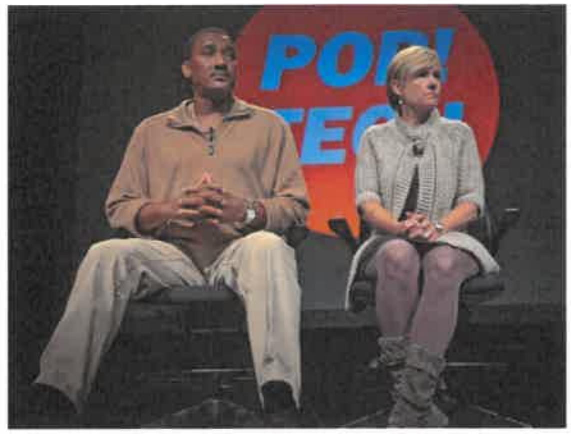
> . . . as one study explained, "[t]he outcomes of empirical studies, reviews, and meta-analyses have converged on the conclusion that the confidence-accuracy relationship for eyewitness identification is weak, with average confidence-accuracy correlations generally estimated between little more than 0 and .29." . . . Another slightly older analysis. . . has suggested a confidence-accuracy correlation of

> only 0.41 for certain types of identifications. . . Importantly, error rates can be high even among the most confident witnesses. Researchers have performed studies that track, in addition to identification accuracy, the subjects' estimates of their confidence in their identifications. In one article reporting results from an empirical study, researchers found that among witnesses who made positive identifications, as many as 40 percent were mistaken, yet they declared themselves to be 90 percent to 100 percent confident in the accuracy of their identifications. . . . This confirms that many witnesses are overconfident in their identification decisions. (American Psychological Association, 2014, pp. 17–18)

Claims like this do not accurately inform the legal system. If these claims of an untrustworthy confidence-accuracy relation had been restricted to specific non-pristine testing conditions that have been shown to compromise the information value of eyewitness confidence or to confidence statements taken later rather than at the time of the initial identification, then they would be defensible claims.

One of the key points we will emphasize is that the only time that confidence is known to be a highly reliable predictor of accuracy is when memory is first tested, before there is much opportunity for memory contamination to occur. An expression of low confidence on that first test is a glaring red flag because it is almost always an indication that the risk of error is high. Instead of being ignored, an initial expression of low confidence should take center stage—overshadowing all other considerations—when a jury's goal is to evaluate the reliability of a suspect ID. If the witness is assumed to be honest, and if the ID was made with low confidence, then it is an unreliable ID. In fact, most of the DNA exonerees who were misidentified by an eyewitness were, at the outset of the investigation, identified with low confidence (Garrett, 2011). It was only later, in court and in front of the jury, that the initial low-confidence ID somehow morphed into a high-confidence ID. If it had been understood that confidence is indicative of accuracy only on an initial memory test (i.e., that on an initial test, low confidence implies low accuracy and high confidence implies higher accuracy), then many of these wrongfully convicted individuals may never have been found guilty in the first place. Or, if prosecutors had understood that low confidence at the initial identification is indicative of a high risk of error, then the innocent suspects in these cases might not have been indicted in the first place. Thus, far from being a problem, *initial* eyewitness confidence is part of the solution to eyewitness-based wrongful convictions (Box 1).

Box 1. Jennifer Thompson's Misidentification of Ronald Cotton



During a trial that was held in 1985, Jennifer Thompson confidently identified Ronald Cotton as the man who had raped her. Cotton was convicted largely on the basis of her testimony, but he was later exonerated by DNA evidence after spending more than 10 years in prison. Long before the trial, however, Thompson's *initial* identification of Cotton from a photo lineup was characterized by a prolonged period of hesitation and indecision that lasted for nearly 5 minutes and ended with a low-confidence verbal identification consisting of the words "I think this is the guy" (Thompson-Cannino, Cotton, & Torneo, 2009, p. 33; Garrett, 2011). However, after confirmatory feedback from the police, Thompson became increasingly confident that Cotton was the rapist. From this perspective, the mistake was to rely on confidence expressed at the time of the trial (after it had become improperly inflated) instead of relying on confidence expressed at the time of the initial ID (before memory contamination had a chance to play a significant role). Indeed, in a very real way, it was the legal system—not Jennifer Thompson—that made the key mistake by ignoring her initial (low) confidence. From this perspective, the time has come to exonerate her, too.

To appreciate how important it is to take into account (not ignore) an initial expression of low confidence by an eyewitness, imagine an eyewitness-identification case involving an innocent suspect that is adjudicated using an approach in which eyewitness confidence is ignored but various factors known to affect eyewitness memory are taken into consideration by a jury. Many of these factors are *estimator variables*—that is, variables that affect memory but are outside of the control of the legal system (Wells, 1978). Some common estimator variables include:

1. Race (cross-race IDs are less accurate than same-race IDs)
2. Exposure duration (brief exposure results in worse memory for the perpetrator than longer exposure)
3. Lighting (poor lighting during the crime results in worse memory for the perpetrator than good lighting)
4. Retention interval (a longer duration between the witnessed crime and the first lineup test results in worse memory for the perpetrator than a shorter duration)
5. Stress (high stress can lead to worse memory for the perpetrator than low stress)
6. Weapon focus (memory for the perpetrator is worse when a weapon is present than when no weapon is present)

For this hypothetical case involving an innocent suspect identified by an eyewitness, assume that all of these factors were favorable. For example, assume it was a same-race ID, exposure duration was long, the lighting was good, the retention interval was short, the witness was not particularly stressed, and no weapon was present. Under such conditions, the jury might reasonably conclude that the eyewitness-identification evidence is reliable and find the innocent suspect guilty. Now imagine that, unbeknownst to the jury, the witness expressed low confidence when the ID was initially made—because the innocent suspect was not a particularly good match to the witness's memory. The evidence we will review shows that such an ID is highly error prone despite the fact that all of the estimator variables are such that one might reasonably conclude otherwise. A low-confidence initial ID ***trumps these good witnessing conditions*** when evaluating the reliability of eyewitness-identification evidence. For that reason, ignoring initial confidence can place innocent suspects at risk.

Whereas low-confidence initial IDs always signal low accuracy—whether the identification procedure was pristine or not—high-confidence IDs on an initial test generally signal high accuracy when pristine testing conditions were used. Thus, initial confidence can serve the cause of justice by protecting the innocent (because initial IDs made with low confidence are untrustworthy) and imperiling the guilty (because initial IDs made with high confidence are trustworthy given appropriate testing conditions). That being the case, it is important to consider what has been learned about the proper way to conduct an eyewitness-identification test.

What Are the Pristine Eyewitness-Identification Procedures?

The error-prone nature of high-confidence eyewitness identifications made in a court of law—after memory has been contaminated—should no longer come as a surprise. All forensic tests—even DNA tests—have the potential to be unreliable if improper testing procedures are used. Proper procedures for obtaining reliable DNA test results were worked out by scientists in the laboratory before they were ever implemented in the forensic domain. As noted earlier, the same cannot be said of eyewitness-identification procedures. Since the 1970s, however, eyewitness-identification researchers have made considerable progress in working out more effective ways of testing eyewitness memory.

A general framework for improving eyewitness-identification procedures was described by Wells and Luus (1990), who proposed the "lineups-as-experiments" analogy. In this analogy, the officer conducting the lineup is like an experimenter; the eyewitnesses are the subjects; instructions to the eyewitnesses can be likened to an experimenter's protocol; the suspect is a stimulus; and the selection of lineup members and the positioning of the suspect in the lineup are part of the design. In addition, police have a hypothesis (e.g., that #4 is the guilty party) and have created a design and procedure to test the hypothesis. The eyewitnesses' choices or identification behaviors constitute the data from which the validity of that hypothesis will be evaluated by police and possibly a prosecutor, judge, and jury.

The idea behind the lineups-as-experiments analogy is that steps that have been taken to enhance the validity of scientific experiments can be applied to police lineup procedures to achieve the same goal. As one example, according to standard laboratory practice, the experimenter is blind to the experimental condition to avoid unconscious biases that might otherwise skew the results in favor of the experimenter's hypothesis. In a police lineup, the "experimenter's" hypothesis is that the suspect is the perpetrator; it therefore stands to reason that the officer administering the lineup should be blind to who the suspect is to avoid unintentionally steering the witness to the suspect. This practice is known as a *double-blind* lineup procedure because neither the lineup administrator nor the witness is told in advance who the suspect in the lineup is. Thus, if the suspect is identified by the witness, one can be more confident that the ID was based on the memory of the witness compared to when a non-blind test is administered.

Another important conceptual distinction to keep in mind is the difference between *system variables* and estimator variables (Wells, 1978). Most eyewitness-identification research has focused on system variables, which are factors affecting the reliability of eyewitness identifications that the criminal justice system can control. As noted earlier, estimator variables are factors that can affect the reliability of an identification but are outside the control of the criminal justice system (e.g., duration of exposure to the perpetrator, the retention interval between the witnessed crime and the first memory test, the presence or absence of a weapon). The main system variable we focus on here concerns how a lineup is administered. Research on lineups has led to a number of recommendations for enhancing the reliability of eyewitness IDs and, critically, for creating the conditions under which confidence is a reliable indicator of accuracy. We review those system-variable recommendations below. Although estimator variables cannot be controlled by the time a crime comes to the attention of the police and thus do not bear on the issue of how to conduct a pristine identification procedure, eyewitness confidence may be an important consideration with respect to those variables as well. Later, following our review of the empirical evidence on the confidence-accuracy relationship, we briefly consider the issue of eyewitness confidence and estimator variables. Here, we consider five system-variable recommendations for the pristine conduct of an eyewitness-identification procedure that were put forward in a white paper of the American Psychology-Law Society and Division 41 of the American Psychological Association (Wells, Small, Penrod, Malpass, Fulero, & Brimacombe, 1998).

Include only one suspect per lineup

A lineup should contain only one suspect, with the remaining persons being known-innocent fillers. The typical recommendation is that a lineup should contain at least five known-innocent fillers (National Institute of Justice, 1999). In the parlance of the lineups-as-experiments analogy, the use of known-innocent fillers can be construed as a method of controlling for guessing. Using an all-suspect lineup, a witness who is prone to simply pick someone will always manage to land on a suspect, and charges might be brought against that person. The dangers of all-suspect lineups have long been documented in the eyewitness-identification literature (Wells & Turtle, 1986). In effect, a lineup that contains only suspects (no fillers) is like a multiple-choice test with no wrong answer. A proper lineup should be constructed in such a way that the witness can "fail" by selecting a filler.

Although fundamental and seemingly elementary, this safeguard against mistaken identification was once commonly violated and is still too often violated today. In fact, in the case of Ronald Cotton and Jennifer Thompson (see Box 1), the photo lineup shown to Thompson was an all-suspect lineup from which she tentatively identified Cotton. This was followed later by a live lineup in which Cotton was the only suspect and the remaining members were fillers. Of course, the actual perpetrator, Bobby Poole, was not in either lineup. The one-suspect recommendation applies under all circumstances. For instance, if there are multiple suspects even though there was only one offender, each suspect should appear in his or her own lineup along with fillers selected for that lineup. If there were multiple offenders, each suspect should still appear in his or her own lineup.

The suspect should not stand out in the lineup

Merely having fillers in a lineup is not in itself a guarantee that they will serve their function of helping to prevent mistaken identifications. Consider, for instance, a case in which the eyewitness described the offender as being a tall, thin male with dark hair and a moustache. Suppose now that the suspect fits this description but some fillers in the lineup are short, others do not have moustaches, and others have light hair. In this case, the suspect will stand out to the witness as being the person who looks most like the offender relative to the other lineup members, regardless of whether the suspect is the actual offender or not. This is the classic idea of a biased lineup. Research shows that placing an innocent suspect who fits the description of the offender in a lineup in which the fillers do not fit the description results in a high rate of mistaken identifications of that person, even when absolute similarity between the innocent person and the offender is only moderate. Moreover, there is evidence for what has been called the *dud effect*, in which adding fillers who look nothing like the perpetrator ("duds") to a lineup increases the confidence with which witnesses choose an innocent person who resembles the perpetrator (Charman, Wells, & Joy, 2011). One way to test whether the fillers are serving their purpose of helping to protect against mistaken identification is to ask whether a non-witness could pick the suspect out from the lineup by merely knowing the description that the eyewitness gave of the offender or by identifying who stands out in the lineup. If the answer is "yes," the fillers are not serving their purpose in the lineup. Indeed, this is the foundation of the "mock witness test" that was developed in the early days of eyewitness-identification research for analyzing the fairness of lineups (Wells, Leippe, & Ostrom, 1979).

Biased lineups are such a severe threat to our ability to rely on the confidence of the witness to infer accuracy that it is important that we give this issue a bit more treatment. One kind of situation that can place an innocent suspect at risk of being mistakenly identified with high confidence is coincidental resemblance between the innocent suspect and the actual perpetrator. Even if all the lineup fillers fit the witness's verbal description of the perpetrator, coincidental resemblance will make an innocent suspect stand out, and empirical studies have shown that unusual resemblance of this type leads to mistaken identifications and high confidence (R. C. L. Lindsay, 1986). We cannot rule out the possibility of coincidental resemblance. But the fact that it is coincidental suggests that it is likely to be extremely rare. In fact, we have found no DNA exoneration case thus far that seems to qualify as having been an example of coincidental resemblance (if by coincidental resemblance we mean that the resemblance was due merely to chance).

On the other hand, unusual resemblance can occur (and has occurred) between an innocent suspect and the perpetrator for reasons other than coincidence. For example, police sometimes use sketch artists or software programs with which witnesses attempt to create a likeness of the perpetrator's face for the purpose of finding possible suspects. In general, if the witness makes a good composite and the composite is then used to find a suspect, the suspect is going to show a strong resemblance to the perpetrator even if the suspect is not the perpetrator. Hence, if the composite is used to find a suspect but the fillers are selected based on the broad verbal description given by the witness, the suspect will stand out (see Box 2).

Box 2. A Striking Resemblance: The Mistaken Identification of Michael McAlister

After spending 29 years in prison for a sexual assault that he did not commit, Michael McAlister was exonerated in 2015. The real perpetrator (on the left) was a serial rapist who bore a striking resemblance to McAlister and the only trial evidence linking McAlister to the attack was the victim's eyewitness identification and testimony. The McAlister case is an example of unusual similarity that, we argue, is not simply a coincidence. McAlister became the suspect in the case based on a facial composite sketch developed with the assistance of the victim witness. If an innocent person becomes a suspect based on their resemblance to a composite sketch (or a surveillance image), there is a heightened risk that the innocent person will have unusual resemblance to the eyewitness's memory of the actual perpetrator. In these cases, the lineup fillers need to be selected based on the fact that they also resemble the composite so as to make sure that the suspect does not stand out in the lineup.

Another way in which an innocent suspect might have unusual similarity to the perpetrator is when surveillance images (e.g., from a convenience store camera) are used to produce a suspect. With the increasing prevalence of electronic surveillance devices in public places, this path to becoming a suspect is likely to be increasingly common. Interestingly, people are quite poor at being able to accurately match a stranger to a surveillance image, even for high-quality images (e.g., see Davis & Valentine, 2009). But the process of using a surveillance image to decide who might be a suspect is rather certain to lead to an individual who is highly similar to the perpetrator, even if the person is innocent. Hence, if an innocent person becomes a suspect and is placed in a lineup based on his or her similarity to a surveillance image, then there is likely to be unusual similarity between the innocent suspect and the eyewitness's memory of the perpetrator, which could lead to a high-confidence (but mistaken) identification.

Notice that this surveillance-image path to high similarity is like the composite example; the high similarity did not occur purely by chance, and therefore it is not coincidental resemblance. And that point is key to solving the problem of unusual similarity when similarity arises from composites or from surveillance images. The solution here is contained in the strategy for selecting fillers for a lineup. Recall that the overall idea for selecting good fillers for a lineup is to make sure that the suspect does not stand out based on what is already known about the perpetrator. For example, if the witness described the perpetrator as being a White male, mid-20s in age, slim build, clean shaven, with short dark hair, and investigators find a suspect with those characteristics, then all of the lineup fillers also need to fit that description. If a composite or surveillance image of the perpetrator was used to find a suspect, however, the composite or surveillance image should trump the verbal description as the criterion for selecting fillers. In other words, if an individual became the suspect based on his or her similarity to a composite or a surveillance image, then the fillers need to also be selected based on their similarity to the composite or surveillance image. Yes, the suspect will still have unusual similarity to the perpetrator even if the suspect is innocent, but so will the fillers. As a result of this strategy for selecting lineup fillers, an innocent suspect should not stand out, thereby controlling the chances of a mistaken identification and false confidence.

Caution that the offender might not be in the lineup

Eyewitnesses often approach lineups with the goal of finding the offender. They should be cautioned that the offender might not be in the lineup because they need to understand that they are not "failing" if they do not choose someone; after all, the correct answer might be "none of the above." In fact, "none of the above" was the correct answer not only in the case of Ronald Cotton and Jennifer Thompson, but also in all the other mistaken-identification cases that have been overturned by DNA testing. The instruction that the perpetrator might not be in the lineup is commonly called the *pre-lineup admonition*.

One concern about the pre-lineup admonition is that it might be undermined by suggestions that occur well before the lineup procedure commences. Quinlivan et al. (2012) found that suggestions to eyewitnesses leading them to believe that the perpetrator would be in the lineup prior to the commencement of the lineup instructions largely canceled the effect of the pre-lineup admonition. This, in turn, increased mistaken identifications in perpetrator-absent lineups and increased the confidence that witnesses had in those mistaken identifications. Consider, for example, an investigator contacting an eyewitness and saying, "We got the guy. We just need for you to come pick him out of a lineup." It seems quite likely that, as Quinlivan et al. found, this suggestion would largely cancel the pre-lineup admonition that would be given later when the formal lineup procedure begins.

Suggestions that occur prior to the commencement of a lineup procedure are concerning because they might be difficult to control. When jurisdictions have adopted pristine eyewitness-identification procedures, those procedures have typically covered only the official commencement of the pre-lineup instructions. Workable solutions to the potential problem of suggestions occurring prior to the initial identification have not been developed; we mention it here simply to raise awareness of it. Although the degree to which it is an actual problem is unknown, it seems reasonable to suppose that it could become more of a problem once the information value of initial eyewitness confidence becomes more widely appreciated. Thus, for the time being, we simply encourage vigilance against this possible contaminating factor.

Use double-blind testing

As noted above, the person who administers a lineup should not know which person in the lineup is the suspect. The use of such double-blind procedures is common in the social and medical sciences. Consider, for instance, the use of placebo control conditions in testing new drugs. Not only is the patient unaware of whether he or she received the drug or a placebo (single-blind), but so are any medical personnel who examine the patients (hence, the term *double-blind*). In this context, "blind" is figurative, not literal. Although the reason for keeping the patient blind as to whether he or she received the drug or a placebo is obvious, the need to keep the tester blind is less obvious.

The reason for keeping the tester blind is to prevent the tester from unintentionally influencing the outcome of the results. The double-blind testing recommendation for lineups does not assume that the tester intends to influence the eyewitness or is even aware of any such influence. This is not an integrity issue. Instead, it is merely an acknowledgment that people in law enforcement, like people in behavioral and medical research, are influenced by their own beliefs and may unknowingly "leak" this information, both verbally and nonverbally, in ways that can influence the person being tested. A vast scientific literature shows that the need for double-blind testing procedures is particularly crucial when there is close face-to-face interaction between the tester and the person being tested (e.g., see Rosenthal & Rubin, 1978).

It should be noted that using a lineup administrator who is blind to the suspect's identity is not the only way to prevent the lineup administrator from influencing the eyewitness. There are other methods, which have been called "blinded" procedures, that prevent the lineup administrator from knowing the position of the suspect in a photo lineup. The U.S. Department of Justice's (1999) guide on eyewitness evidence, for example, describes a folder or envelope "shuffle" method to prevent the officer from knowing which photo the witness is viewing. The shuffle method can be used for both simultaneous and sequential lineups, as it was in the blinded condition of a recent police department field study (Wixted et al., 2016). Alternatively, photo lineups can be administered using laptop computers that shuffle the order of the array, with the screen kept out of view of the lineup administrator.

Collect a confidence statement at the time of the identification

At the time an eyewitness makes an identification, a statement should be obtained from the eyewitness indicating how confident he or she is that the person identified is the offender. Of course, this assumes double-blind testing: The statement should be obtained by a lineup administrator who does not know which lineup member is the suspect. It is this initial confidence statement—and only this confidence statement—that is known to be a reliable indicator of accuracy. As we note in the next section, later statements of confidence by the eyewitness may not be reliable indicators of accuracy because confidence is malleable as a function of later events.

Additional Notes on Concerns About Contamination of Confidence

Before we discuss the issue of measuring the confidence-accuracy relation, we offer a deeper discussion of factors that can contaminate witness confidence and threaten its relation to accuracy. This discussion can help produce a better understanding of the five recommendations for pristine procedures that were discussed above as they relate to witness confidence.

The confidence that people have in a memory is malleable. Studies show that simply imagining that some childhood event happened (when in fact it did not) can lead people to develop false confidence that they remember the fictitious event actually happening (Garry, Manning, Loftus, & Sherman, 1996). In the case of eyewitness identification, both the anticipation by eyewitnesses that they will later be cross-examined about their identification and the encouragement to prepare themselves for cross-examination have been shown to inflate witnesses' confidence (e.g., Wells, Ferguson, & Lindsay, 1981). Presumably, this confidence inflation occurs because witnesses rehearse the event in preparation for cross-examination, which makes the memory more vivid and fluently retrieved and thereby makes it seem more true, even if it is a false memory. Again, however, our thesis about the diagnosticity of confidence applies only to the initial confidence of the witness at the time of identification, not to later feelings of confidence that might be the product of post-identification contamination.

Perhaps the biggest threat to our ability to rely on confidence in eyewitness identification occurs when witnesses receive post-identification feedback that suggests they made an accurate identification (Wells & Bradfield, 1998). There is now a large body of eyewitness-identification studies showing that a simple comment to an eyewitness who has made a mistaken identification (e.g., "Good, you identified the suspect") can lead to immediate strong inflation of the witness's confidence. The effect of post-identification feedback is large. A recent meta-analysis of post-identification feedback studies showed that the eyewitnesses' confidence in their mistaken identifications was inflated by approximately a full standard deviation following such a comment (Steblay, Wells, & Douglass, 2014). The post-identification feedback effect is more muted for accurate eyewitness identifications, which means that confirmatory post-identification feedback actually harms the relation between accuracy and confidence (Charman & Wells, 2012).

There is a provocative and important twist to the post-identification feedback effect. Specifically, in post-identification feedback experiments, the question asked of witnesses is "How confident were you *at the time of the identification?*" Whereas few might be surprised that witnesses' post-identification confidence is inflated by confirmatory post-identification feedback, these studies measure the witnesses' retrospective confidence (not current confidence) by asking them to report how confident they recall having been at the time of the identification (before they received the feedback). So, post-identification feedback not only affects current confidence but also distorts eyewitnesses' recall for how confident they were at an earlier time. In fact, in a *New York Times* op-ed in 2000, Jennifer Thompson had this to say about her initial mistaken ID of Ronald Cotton: "Several days later, looking at a series of police photos, I identified my attacker. I knew this was the man. I was completely confident. I was sure" (Thompson, 2000). In truth, Jennifer Thompson was not completely confident at the time: Her initial ID was made with low confidence. However, feedback that she received at a later time led her to erroneously recall having been sure from the outset.

Interestingly, when witnesses were asked if post-identification feedback might have influenced how they answered the confidence question, most said "no," yet those who said "no" were no less influenced than were those who said "yes" (Wells & Bradfield, 1998). Moreover, post-identification feedback produces this same type of distortion not just for retrospective confidence but also for other testimony-relevant self-reports of eyewitnesses, such as reports of how much attention they paid at the time of witnessing and how good their view was of the perpetrator (see Steblay et al., 2014, for a meta-analysis of all these measures).

Experimental evidence indicates that lineup administrators' own expectations are likely to influence the confidence of the witness even when the lineup administrators are given an objective script to follow and are instructed to not deviate from that script. Garrioch and Brimacombe (2001) randomly assigned people to the role of witness or lineup administrator and then randomly assigned lineup administrators to a condition in which they were led to believe that the perpetrator was in a particular position of the lineup or a condition in which the lineup administrators were told nothing about the perpetrator position in the lineup. In reality, the perpetrator was never in the lineup. But when witnesses chose the lineup member who the lineup administrator had been led to believe was the perpetrator, the witnesses reported being much more confident than when the lineup administrator had no expectations about which person was the perpetrator. Videotapes of the lineup administrators' behaviors showed different patterns of post-identification nonverbal or paralinguistic behaviors as a function of lineup administrators' expectations about which lineup member was the perpetrator. These influences of the lineup administrators' expectations on the confidence of the witnesses occurred despite the fact that there were no incentives or other motivations on the part of the lineup administrators. Furthermore, 100% of the lineup administrators indicated that they believed they did not provide any post-identification feedback, and 95% of the witnesses believed they did not receive any post-identification feedback.

So, post-identification feedback appears to be a pernicious problem. Fortunately, we have long known the solution for preventing the contamination of post-identification feedback, namely the double-blind lineup procedure that eyewitness researchers have been proposing for over 25 years (Wells & Luus, 1990). In fact, one of the primary reasons for double-blind lineup testing is to prevent the lineup administrator from giving inadvertent feedback that could distort the confidence of the witness. The simple beauty of the double-blind lineup procedure is that the lineup administrator does not know if the witness picked a known-innocent filler or picked the suspect in the case. That same double-blind lineup administrator can then secure a confidence statement from the witness prior to any opportunity for the witness to be given feedback about whether the identified person was the suspect or was a lineup filler (Wells & Bradfield, 1999).

In addition to conducting the lineup using a double-blind procedure, eyewitness-identification researchers have long advocated videotaping the entire eyewitness-identification procedure (e.g., Kassin, 1998). And the idea of videotaping all identification procedures was recently endorsed by a committee of the National Academy of Sciences (National Research Council, 2014). The initial confidence statement is then a matter of record, and it is that initial confidence statement, not later confidence statements, that prosecutors and courts should rely upon. If the case reaches trial, juries should use only this initial confidence statement for assessing the reliability of the identification.

Of course, having a pristine assessment of witness confidence at the time of the identification does not prevent witnesses from undergoing confidence inflation later and perhaps being quite positive at trial. But that is why we emphasize so strongly that the reliability of confidence statements must be based on the eyewitness's initial confidence, not later claims of confidence. And this is where courts have commonly made a serious mistake. Courts routinely permit witnesses to state their confidence at pretrial hearings or at trial, well after they might have undergone serious confidence inflation from repeated identifications, coaching, confirmatory feedback, and so on. The confidence of the witness at the time of a preliminary hearing or at trial is not a pristine assessment of confidence.

Interestingly, the U.S. Supreme Court's guiding ruling on eyewitness identification, which is now nearly 40 years old, urged lower courts to consider the confidence that the eyewitness had *at the time of the identification* in evaluating the reliability of an eyewitness identification (*Manson v. Braithwaite*, 1977). What can be done if the lineup administrator failed to secure a confidence statement from the witness at the time of the identification? Some courts might be tempted to simply ask the witness to cast his or her mind back to the lineup and recall how confident he or she was at the time of the identification. But, as the literature on the post-identification feedback effect shows, witnesses do not accurately recall their initial uncertainty if confidence inflation has occurred as a result of contaminating influences, and instead recall having been confident all along. There is no substitute for taking a confidence statement at the time of the identification.

It is also important to keep in mind that our claims about the reliability of confidence as an indicator of accuracy in eyewitness identification apply only to cases in which the eyewitness-identification test procedures were pristine (Box 3). Unfortunately, at this point, not all jurisdictions in the United States collect a confidence statement at the time of the identification, and when they do, they often do not use a double-blind procedure. Indeed, as recently as 2001, there was no jurisdiction in the United States that used double-blind lineup procedures. Fortunately, efforts by eyewitness-identification researchers, in partnership with the Innocence Project, local and state-level reform commissions, and other policymakers, have managed to facilitate reforms on eyewitness-identification procedures in a growing number of jurisdictions in the United States. As of the time of this writing, for example, state laws have been passed by legislators that require double-blind lineup administration in Connecticut, Colorado, Kansas, Illinois, Maryland, North Carolina, Ohio, and Vermont. Additional states have used other mechanisms to force the use of double-blind lineup administration. New Jersey, for example, requires double-blind lineup administration via a plenary mandate from the attorney general of New Jersey. Oregon's Supreme Court issued a decision (*State v. Lawson*, 2012) that largely makes double-blind lineup procedures necessary in Oregon. In addition, the states of Texas, Rhode Island, Wisconsin, and West Virginia have achieved substantial compliance for using double-blind lineup procedures through a combination of laws and influential task-force recommendations. In addition, individual jurisdictions such as Suffolk County, Massachusetts (Boston and surrounding areas), Santa Clara County, California (including San Jose and Palo Alto), Minneapolis, Minnesota, and many other large and small jurisdictions have made eyewitness-identification reforms that include the requirement of double-blind lineup administration. At the time of this writing, numerous other states are considering requiring double-blind lineup administration.

Box 3. Pristine Lineup Conditions

1. Include only one suspect per lineup
2. The suspect should not stand out in the lineup
3. Caution that the offender might not be in the lineup
4. Use double-blind testing
5. Collect a confidence statement at the time of the identification

Because there remain many jurisdictions that have not yet adopted pristine eyewitness-identification testing procedures, it is important that we emphasize a caveat to our primary thesis. Specifically, our claim regarding the high diagnosticity of eyewitness identifications made with high confidence does not extend without qualification to those jurisdictions that have not yet made reforms to ensure pristine procedures. For example, as we will show later, a high-confidence ID made from an unfair lineup is considerably more error prone than a high-confidence ID made from a pristine lineup. A similar risk of error may occur when eyewitness-identification procedures depart from the other recommended procedures as well, though detailed investigations into their effect on high-confidence accuracy have not been performed. Nevertheless, it seems safe to say that prosecutors and defense attorneys are likely to debate the reliability of a suspect ID whenever the procedures summarized in Box 3 have not been followed, and for good reason. For example, if a lineup was administered in a non-blind fashion, the question will inevitably arise as to whether the lineup administrator unintentionally influenced the identification made by the witness and the confidence of the witness, which research shows is a real possibility (e.g., Garrioch & Brimacombe, 2001). As noted in the National Research Council (2014) report, "The use of double-blind procedures will eliminate a line of cross-examination of officers in court" (p. 107). The same argument can be made for each of the practices listed in Box 3.

Whereas the information value of a high-confidence ID may be called into question whenever non-pristine testing conditions are used, the information value of a low-confidence ID is never open to question. No matter how good or how bad the eyewitness-identification procedure is, a low-confidence ID implies that the ID is error prone. As noted above, if an identification was made in a jurisdiction that has not adopted pristine testing conditions, the defense and the prosecution may end up debating in court about whether or not the testing procedure was good enough. However, that debate is rendered moot if it is known that the eyewitness made an initial good-faith ID with low confidence. Such an ID is error prone, even under pristine testing conditions.

Returning to the main point, if the pristine conditions listed here (summarized in Box 3) are followed, then a low-confidence ID implies low accuracy, and a high-confidence ID implies high accuracy. Although eyewitness-identification research conducted over the last 20 years has shown this to be true, our understanding of this issue emerged rather gradually, which may help to explain why it is not more widely understood within the legal system. We turn now to a consideration of the eyewitness confidence-accuracy literature, beginning with a review of the methods used to measure the confidence-accuracy relationship (a key part of the story). We emphasize that except where noted, the studies we consider were carried out using the pristine testing conditions summarized above. How reliable is an ID made under those conditions, according to what we have learned over the last 20 years? To answer that question, we first have to consider which approach to measuring the confidence-accuracy relation most accurately conveys the information sought by judges and juries.

Measuring the Eyewitness Confidence-Accuracy Relationship

The data used to investigate the confidence-accuracy relationship for eyewitness identification come mostly from forensically relevant lab studies in which participants become witnesses to a mock crime (e.g., by watching a live enactment or a video of someone committing a crime, such as snatching a purse, planting a bomb, or robbing someone at an ATM) and are later shown a lineup in which the perpetrator (the target) is either present or absent. A target-present lineup includes the perpetrator along with (usually five or seven) similar fillers; a target-absent lineup is the same except that the perpetrator is replaced by another similar filler, as illustrated in Figure 1. In some studies, the individual depicted in the replacement photo serves the role of the designated "innocent suspect." In other studies, no one in the target-absent lineup is designated to serve the role of an innocent suspect, so the risk to an innocent suspect is calculated by dividing the number of identifications in the target-absent lineup by the number of fillers (thereby assuming a perfectly fair lineup). When presented with a target-present or target-absent lineup, a witness in a mock-crime study first makes a decision—which consists of identifying the suspect, identifying a filler, or rejecting the lineup (i.e., saying that the perpetrator is not there)—and then supplies a confidence rating associated with that decision. A correct response consists of (a) a suspect

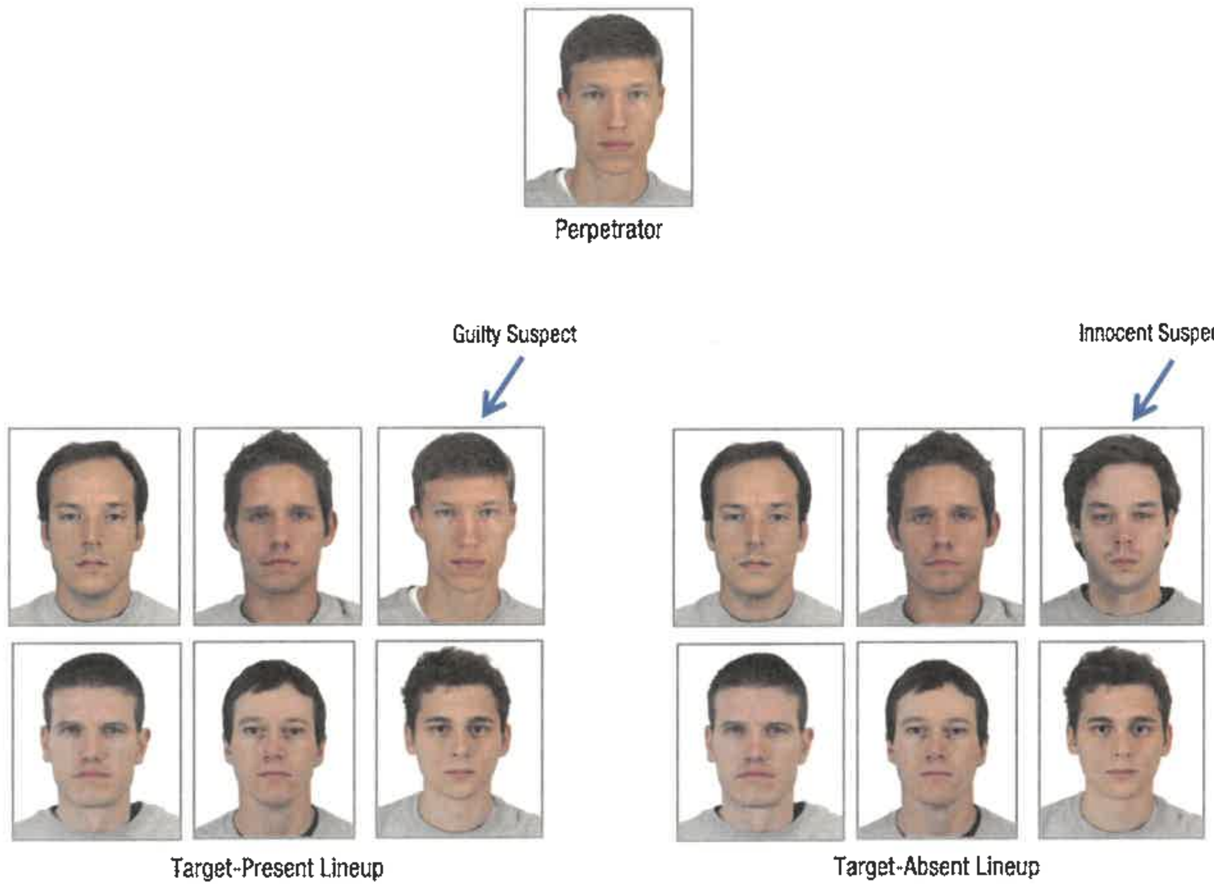


Fig. 1. Example lineups used in mock-crime studies. (Images drawn from the Chicago Face Database; Ma, Correll, & Wittenbrink, 2015.)

ID from a target-present lineup or (b) the rejection of a target-absent lineup, whereas an incorrect response consists of (a) a suspect ID from a target-absent lineup (if there is a designated innocent suspect), (b) a filler ID from either type of lineup, or (c) the rejection of a target-present lineup.

Our appreciation of the information value of confidence has grown considerably in recent years, partly as a result of methodological changes in the way that researchers measure the confidence-accuracy relationship. Prior research on the issue can be divided into three phases according to the measure that was used. In Phase 1, the point-biserial correlation coefficient was the preferred measure. In Phase 2, calibration curves were more commonly used. In Phase 3, confidence-accuracy characteristic (CAC) analysis (Mickes, 2015) and closely related but more complete Bayesian analyses (Wells, Yang, & Smalarz, 2015) have been used. An argument we will advance is that only the measures used in Phase 3 directly address questions of interest to the legal system.

The point-biserial correlation coefficient

In Phase 1, the relationship between confidence and accuracy was measured by computing the standard Pearson *r* correlation coefficient between the accuracy of a response (e.g., coded as 0 or 1) and the corresponding confidence rating (e.g., measured using a 5-point scale from *just guessing* to *very sure that is the person*). Because accuracy is coded as a dichotomous variable, the Pearson *r* in this case is known as a point-biserial correlation coefficient. Using this measure, much of the initial research examining eyewitness certainty suggested that certainty was largely unrelated to identification accuracy (e.g., Clifford & Scott, 1978; Deffenbacher, Brown, & Sturgill, 1978; Leippe, Wells, & Ostrom, 1978), with correlation coefficients generally falling into the .00 to .20 range.

In these studies, all of the data were bundled together for the analysis, whether the eyewitness made a suspect ID, a filler ID, or a non-ID. In a later meta-analysis, Sporer, Penrod, Read, and Cutler (1995) found that the relationship was noticeably stronger—about .41—when the analysis was limited to only those who made an ID from a lineup (i.e., when the analysis was limited to "choosers" who identified a suspect or a filler). Limiting the analysis to choosers is reasonable because only witnesses who choose someone end up testifying in court against the person they identified. This move—separating choosers (those who make suspect IDs and filler IDs) from non-choosers (those who make a non-ID)—foreshadowed a later move that we will argue is critical, namely separating choosers who make suspect IDs from choosers who make filler IDs. All three decision outcomes (suspect IDs, filler IDs, and non-IDs) need to be assessed for the independent information they provide about whether or not the lineup contains a guilty suspect. Nevertheless, Sporer et al.'s separation of choosers from non-choosers led to an important advance in our understanding of the confidence-accuracy relationship.

The novel message from Sporer et al. (1995) was that confidence is a more reliable indicator of accuracy for choosers than had been previously assumed. At .41, the correlation for choosers was clearly too large to argue that eyewitness confidence should be disregarded. Nevertheless, over the years, the interpretation of the Sporer et al. (1995) meta-analysis has generally drifted in the negative direction, as if the message were actually the opposite. For example, Reinitz, Seguin, Peria, and Loftus (2012) said, "It is well known that confidence is often a relatively poor predictor of accuracy (e.g., Bothwell, Deffenbacher, & Brigham, 1987; Sporer et al., 1995)" (p. 1089). Buratti and Allwood (2012) noted that "although many witnesses may feel confident about their identification, the relation between identification confidence and the correctness of the identification is weak (Brewer & Wells, 2011; Sporer et al., 1995)" (p. 590). Neal, Christiansen, Bornstein, and Robicheaux (2012) pointed out that "contrary to jurors' beliefs, eyewitness confidence is not a strong indicator of accuracy (Penrod & Cutler, 1995; Sporer et al., 1995)" (p. 50). And Wilson, Hugenberg, and Bernstein (2013) recently maintained that "one surprising lesson that psychologists have learned about memory is that the confidence of an eyewitness is only weakly related to their recognition accuracy (p. 98; see Sporer et al., 1995, for a review)."

It seems fair to say that these characterizations do not accurately convey what the Sporer et al. (1995) meta-analysis actually found. What Sporer et al. actually found was that, for choosers, the confidence-accuracy relationship is surprisingly strong. They also emphasized the fact that later events can inflate an eyewitness's confidence, obviously without increasing the accuracy of the initial ID. Some of the post-ID factors that can inflate confidence include hearing that other witnesses have identified the same suspect (Luus & Wells, 1994), being exposed to the identified suspect again (Brown, Deffenbacher, & Sturgill, 1977), and receiving encouraging feedback from police about the accuracy of the ID (Wells & Bradfield, 1998). However, for an *initial* ID made from a pristine lineup, the Sporer et al. (1995) meta-analysis showed that initial confidence is a reasonably good indicator of accuracy. In fact, the point-biserial correlation coefficient is a standard effect-size statistic (e.g., Rosnow, Rosenthal, & Rubin, 2000), and a value of .41 falls between the conventional definitions for medium (.30) and large (.50) effects (Cohen, 1988).

Shortly after Sporer et al.'s (1995) meta-analysis was published, the argument was made that even their upgraded assessment of the confidence-accuracy relationship was, if anything, an understatement. Juslin et al. (1996) showed that the magnitude of the point-biserial correlation can be low even when the relationship between confidence and accuracy exhibits perfect calibration. Perfect calibration exists when the level of confidence expressed by an eyewitness corresponds exactly to the percentage of eyewitnesses who are correct when they express that level of confidence. For example, under perfect calibration, witnesses who express 60% confidence in an ID are correct 60% of the time, and witnesses who express 80% confidence in an ID are correct 80% of the time. Even though the relationship between confidence and accuracy could not possibly be stronger than that, Juslin et al. showed that the point-biserial correlation could be low or high, depending on how responses are distributed across the confidence categories. In Appendix A, we provide a concrete example illustrating how this could be. The key point is that the .41 correlation coefficient for choosers is potentially compatible with a *very* strong confidence-accuracy relationship.

These considerations suggest that the point-biserial correlation coefficient is not the best statistic to use when trying to inform the legal system about the utility of eyewitness confidence. Note that this is not a criticism of the statistic itself. The point-biserial correlation coefficient is a perfectly valid effect-size statistic when used in conjunction with certain statistical tests, such as a *t* test (Rosnow et al., 2000). For example, in eyewitness-identification studies, one might ask whether the average level of confidence is higher for correct IDs than for incorrect IDs. This would be the appropriate way to analyze the data if you knew, for each eyewitness, whether his or her ID was correct or incorrect and you wanted to estimate his or her likely level of confidence. In fact, this is how the data were plotted in Figure 1 of Sporer et al.'s (1995) seminal article, and the corresponding point-biserial correlation coefficient of .41 indicates a moderate-to-large average effect size. Yet this is not the question of interest to the legal system, because in actual practice, the situation is reversed: An eyewitness provides a confidence rating associated with an ID (this is the predictor variable, which is not averaged), and the legal system wants to make the best estimate as to the likely accuracy of that ID (this is the dependent variable, and it equals the average level of accuracy associated with each level of confidence that an eyewitness might express).

This logic suggests, as Juslin et al. (1996) pointed out, that plotting average accuracy (on the *y*-axis, as the dependent measure) versus different levels of confidence (on the *x*-axis, as the independent measure) is a more informative way to analyze the data. When plotted this way, the data come closer to providing an answer to the question asked by judges and juries trying to evaluate the reliability of an eyewitness. Their question is: Given that an eyewitness has a particular level of confidence in his or her ID of a suspect, how accurate is that ID likely to be? With regard to that question, a calibration curve provides much more relevant information than a correlation coefficient. Once this fact was understood, Phase 2 was ushered in as eyewitness-identification researchers began to measure the confidence-accuracy relationship by plotting calibration curves.

Calibration analysis

Following Sporer et al. (1995), calibration analyses are also typically performed separately on choosers (those who make a suspect ID or a filler ID) and non-choosers (those who make a non-ID decision). A calibration analysis can be performed whenever a confidence rating scale ranging from 0 to 100 is used. It is important to be clear about the exact computational formula used to compute calibration, so we consider the formula below. In the notation we use here, ***nFID*** stands for "number of filler IDs" and ***nSID*** stands for "number of suspect IDs." We also attach subscripts to these symbols, such as *TP*, which denotes target-present lineups, and *TA*, which denotes target-absent lineups. Thus, $nSID_{TP}$ means "number of suspect IDs from target-present lineups." Finally, we add the subscript *c*, which represents the confidence expressed by the witness. Thus, $nSID_{TP\text{-}c}$ means "number of suspect IDs from target-present lineups with confidence *c*," where *c* might be 90% to 100% confident.

Basically, in a calibration analysis of choosers, the percentage-correct accuracy score for a given level of confidence, *c*, is equal to 100 multiplied by the number of (correct) suspect IDs from target-present lineups made with confidence level *c* ($nSID_{TP\text{-}c}$) and divided by the total number of IDs (to suspects and fillers alike) made with confidence level *c*. Many calibration studies have used a target-absent lineup that does not have a designated innocent suspect, so the number of incorrect IDs consists of the number of filler IDs made from target-present lineups with confidence level *c* ($nFID_{TP\text{-}c}$) plus the number of filler IDs made from target-absent lineups with confidence level *c* ($nFID_{TA\text{-}c}$). Thus, for confidence level *c*, calibration equals $100 \times (nSID_{TP\text{-}c}) / (nSID_{TP\text{-}c} + nFID_{TP\text{-}c} + nFID_{TA\text{-}c})$. In practice, $nFID_{TP\text{-}c}$ is often excluded from the denominator, but the results tend to be similar either way.

Calibration studies typically find a strong relationship between confidence and accuracy when (a) the analysis is limited to choosers, (b) the witnesses are adults, (c) the lineups are fair, and (d) the confidence ratings are taken immediately after the ID is made (e.g., Brewer & Palmer, 2010). That is, they find a strong relationship between confidence and accuracy using pristine eyewitness-identification procedures that were also used in previous studies measuring the relationship using the point-biserial correlation coefficient. As an example, Figure 2a presents a calibration curve taken from Brewer and Wells (2006). As we will see, the results shown in Figure 2a are fairly typical of calibration studies, and they show that low-confidence IDs (*c* = 0%–20%) are associated with low accuracy (26.6% correct), whereas high-confidence IDs (*c* = 90%–100%) are associated with much higher accuracy (84.9% correct). It seems difficult to characterize the results shown in Figure 2a as indicating anything other than a very strong confidence-accuracy relationship for choosers. This is true even though the overall point-biserial correlation coefficient in this study was low (.32 for identifications of the thief in the video and .36 for identifications of the waiter in the video). These findings underscore the fact that the confidence-accuracy correlation can be low even when the confidence-accuracy relationship is strong. Note that the story for non-choosers is different. For them, the confidence-accuracy relationship is noticeably weaker (Fig. 2b), which is a conclusion that also corresponds to work using the point-biserial correlation coefficient (Sporer et al., 1995).

These results, like the point-biserial results discussed above, correspond to the 50% base rate of target-present lineups used in that study. As we discuss in more detail later, real police lineups may contain a guilty suspect less than 50% of the time. In such cases, the accuracy rates for choosers would be correspondingly lower than the values shown in Figure 2a, and the accuracy rates for non-choosers would be correspondingly higher than the values shown in Figure 2b. Nevertheless, the basic story would not change: For choosers, the confidence-accuracy relationship is strong, and for non-choosers it is considerably weaker.

Although the results in Figure 2a reflect a strong confidence-accuracy relationship, it also seems fair to say—and it often is said—that witnesses who express high

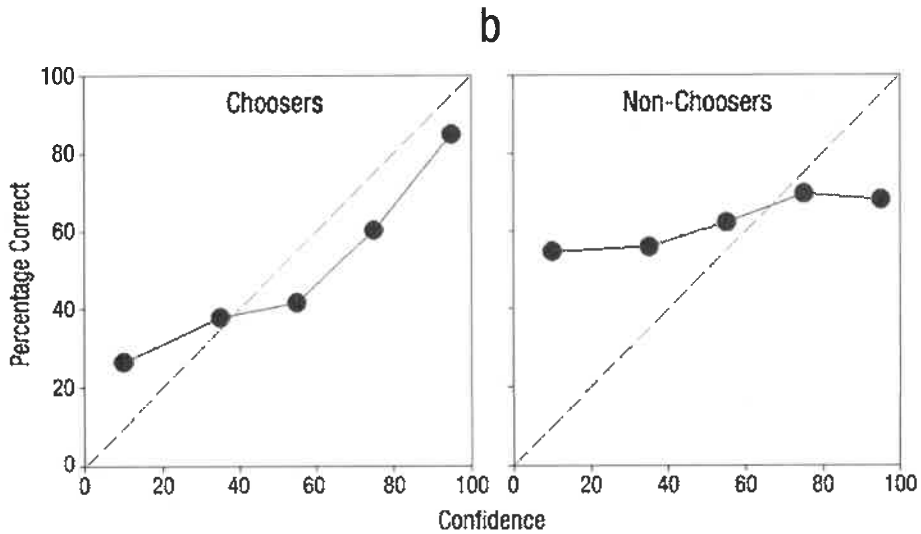


Fig. 2. Calibration data from Brewer and Wells (2006) for choosers (a) and non-choosers (b). The data were pooled across witnesses' identifications of either of the two targets who appeared in a mock-crime video (a thief and a waiter). The dashed line represents perfect calibration.

confidence, such as 90% to 100% confidence, are overconfident because their corresponding accuracy is typically lower than 90% (e.g., Lampinen, Neuschatz, & Cling, 2012; Leach, Cutler, & Van Wallendael, 2009; Valentine & Davis, 2015). Leach et al. (2009) put it this way:

> Given the modest correlation between confidence and identification accuracy, the tendency for witnesses to be overconfident in their decisions (Brewer & Wells, 2006), and the factors that further suppress the confidence-accuracy relation, confidence is of questionable utility in the assessment of eyewitness identification accuracy. (p. 161)

However, this pessimistic assessment seems premature, because the data have still not yet been analyzed in a way that most directly addresses the question of interest to judges and juries tasked with assessing the reliability of an initial eyewitness ID made with a particular level of confidence. In the courtroom, the question of interest is as follows: What does confidence tell us about the reliability of an eyewitness who has identified a *suspect?* The answer to this question is provided by limiting the analysis not just to choosers but to choosers who identify a suspect—just as the legal system limits its consideration to choosers who identify a suspect by referring identified suspects (but not identified fillers) for prosecution.

We refer to the dependent variable in an analysis that excludes filler IDs as *suspect-ID accuracy*, and we refer to a plot of suspect-ID accuracy versus confidence as a CAC analysis to distinguish it from the closely related calibration plot (Mickes, 2015). Unlike a calibration curve, a CAC plot provides the information that judges and juries want to know when they are trying to assess the reliability of an eyewitness who identified a suspect from a lineup.

Once again, our argument should not be construed as a criticism of the calibration statistic. A calibration curve is a perfectly appropriate way to represent the relevant data when the question concerns the confidence-accuracy relationship from the witness's perspective. In a calibration study, witnesses are instructed to choose a confidence rating of 80% (for example) when they believe they would be correct 80% of the time. From the witness's perspective, a correct ID consists of choosing a suspect from a target-present lineup, whereas an error consists of choosing a suspect from a target-absent lineup or choosing a filler from either type of lineup. Thus, a witness presumably interprets the instruction to mean that a confidence rating of 80% should consist of correct responses (suspect IDs from target-present lineups) 80% of the time and errors (suspect IDs from target-absent lineups and filler IDs) 20% of the time. A calibration curve appropriately shows the relationship between what a witness believes about his or her performance and what that performance is actually like.

However, the legal system is concerned with a different issue, because if the eyewitness picked a filler, we already know that the witness did not pick the perpetrator. So, the forensically relevant question is this: Given that the eyewitness picked the suspect with a particular level of confidence, how likely is it that the suspect is guilty? The answer to that question is provided by a CAC plot in which the dependent measure is suspect-ID accuracy. We next describe how to compute suspect-ID accuracy, and then we reanalyze and plot the published data in terms of CAC analysis. We then use representative calibration data and reanalyze those results using the more detailed Bayesian analysis described by Wells et al. (2015). This analysis shows suspect-ID accuracy across the full range of base rates of target-present lineups (instead of limiting the analysis to the 50% base rate typically used in studies, as CAC analysis does). Using the same basic approach, we also consider a topic that is only rarely considered: What is the information value of a filler ID or a non-ID? These decision outcomes also bear on the likelihood that the suspect in the lineup is guilty, but the information they provide points in the opposite direction than that provided by a suspect ID, in that they are both probative of innocence (Wells et al., 2015). That fact is another reason suspect IDs and filler IDs should not be bundled together when the goal is to inform the legal system. They should not be bundled together because they provide independent (and opposing) information about the likelihood that the suspect in the lineup is guilty.

Confidence-accuracy characteristic analysis

Suspect-ID accuracy is based on the number of suspect IDs from target-present lineups (guilty-suspect IDs) made with confidence level *c*, $nSID_{TP\text{-}c}$, and the number of suspect IDs from target-absent lineups (innocent-suspect IDs) made with the same confidence level, $nSID_{TA\text{-}c}$. More specifically, suspect-ID accuracy is equal to 100% × $nSID_{TP\text{-}c}$ / ($nSID_{TP\text{-}c}$ + $nSID_{TA\text{-}c}$). Unlike in a real police lineup involving an innocent suspect, in a lab study there is no obvious person to use as the innocent suspect. In other words, there is obviously no one in the target-absent lineup who is suspected of having committed the crime depicted in the mock-crime video (because the experimenter selected the perpetrator in the video and so already knows who he is). How, then, does one compute the number of innocent-suspect IDs? Using one reasonable approach, the innocent suspect in a target-absent lineup is simply a designated filler, usually the filler that was used to replace the perpetrator's photo (as in Fig. 1). This approach is arguably the most logical approach because only the suspect differs across target-present and target-absent lineups (the other fillers are held constant). It also has the advantage of making it easy to count not only the number of guilty-suspect IDs that are made from target-present lineups but also the number of innocent-suspect IDs that are made from target-absent lineups. In that case, computing suspect-ID accuracy for each level of confidence is entirely straightforward.

If the replacement photo is not designated as the innocent suspect in target-absent lineups, then $nSID_{TA}$ can instead be estimated from the number of filler IDs from target-absent lineups divided by lineup size (*n*). In that case, suspect-ID accuracy would be given by 100% × $nSID_{TP\text{-}c}$ / ($nSID_{TP\text{-}c}$ + $\sim nSID_{TA\text{-}c}$), where $\sim nSID_{TA\text{-}c}$ = ($nFID_{TA\text{-}c}$ / *n*). For most of the studies we will review, the number of IDs of the replacement photo was not reported, so this approach to estimating the false-ID rate was the only option available. In the long run, these two approaches to computing the false-ID rate (namely, designating an innocent suspect vs. counting all target-absent filler IDs and dividing by lineup size) will yield the same average results so long as all of the fillers—including the one that replaces the perpetrator—are selected using the same decision rule, such as the rule that fillers must match the description of the perpetrator. For any particular study, however, the two approaches can yield different results. For example, by chance, half the time, the replacement photo (i.e., the natural choice to serve as the innocent suspect) will be a more attractive option than the average filler in the lineup, and 1/*n* of the time it will, by chance, be the most attractive option. In these studies, dividing the target-absent filler-ID rate by *n* will underestimate the false-ID rate, making suspect-ID accuracy at each level of confidence look better than it would look if the replacement photo had been designated as the innocent suspect. On the other hand, the other half of the time, the replacement photo will be a less attractive option than the average filler in the lineup, and 1/*n* of the time it will be the least attractive option. In these studies, dividing the target-absent filler-ID rate by *n* will overestimate the false-ID rate, making accuracy at each level of confidence look worse than it would if the replacement photo had been designated as the innocent suspect. Thus, our conclusions about the confidence-accuracy relationship will be based on what studies suggest in the aggregate, not on what any particular study suggests. Nevertheless, in light of these considerations, we believe that researchers should report the frequency with which each target-absent lineup member was identified. In any given study, it might be the case that, by chance, the replacement filler was chosen more often than the other fillers. If so, a conclusion derived from that study alone would apply more to unfair lineups than to fair lineups. When target-absent filler-ID rates for each lineup member are not reported, there is no way to tell if this is a problem or not.

A third approach to designating an innocent suspect in a target-absent lineup is problematic if the goal is to measure the confidence-accuracy relationship under pristine testing conditions. Using this third approach, the innocent suspect is defined to be the filler in the target-absent lineup who most resembles the perpetrator. The innocent suspect stands out in that sense and will therefore be chosen more often than the other fillers. For example, in a recent study, Sučić, Tokić, and Ivešić (2015) first selected a set of six fillers who matched the description of the target in the target-present lineup and then selected the one who would serve as the designated innocent suspect in the following way: "The six photographs that were selected were top ranked for photograph similarity and the match to modal description, and the highest ranked photograph was used as the designated innocent suspect (suspect replacement) in a [target-absent] lineup" (p. 802). In other words, the designated innocent suspect in this study was chosen in such a way as to ensure that it would stand out in the target-absent lineup (i.e., by design, this was an unfair lineup). As noted earlier, in an ideal lineup, the suspect does *not* stand out, and if the police made lineups following this approach, those lineups would be unfair. We will separately review studies that used this approach, and we will see that it has a profoundly negative effect on the confidence-accuracy relationship even for an otherwise pristine identification procedure. However, the bulk of our review will consist of a reanalysis of studies that used lineups in which the replacement photo in the target-absent lineup was not selected using a different decision rule than the other fillers. For these studies, the number of innocent-suspect IDs was estimated by counting all filler IDs from target-absent lineups and dividing by lineup size. Again, for any single study, this approach to estimating the false-suspect-ID rate could mask the fact that the target-absent lineup was, by chance, biased toward or away from the replacement photo (i.e., toward or away from the photo that would most logically serve as the innocent suspect). Our conclusions about the relationship between confidence and accuracy are not based on any single study but are instead based on the aggregate results from many studies.

It is important to emphasize that the suspect-ID accuracy measure in CAC analysis is not another measure of calibration. As described in more detail below, random chance accuracy for suspect IDs is typically 50% correct, not 0% correct. Thus, if a 0-to-100 confidence scale is used, one would not expect to see suspect-ID accuracy match the level of confidence at the low end of the scale. In a CAC analysis, the question is not how well confidence and accuracy match; instead, the question is simply this: How does suspect-ID accuracy vary as confidence ranges from low to high? Because that is the question, CAC analysis can be carried out using any monotonic confidence rating scale (unlike a calibration analysis, which requires a 0-to-100 scale). In point of fact, very few police departments use 0-to-100 scales to assess initial confidence, so calibration is not often at issue in the legal system (although it is of interest in laboratory studies).

Suspect-ID accuracy is what judges and juries want to know when trying to evaluate the reliability of an eyewitness identification: Given that the suspect in this trial has been identified by an eyewitness with a particular level of confidence, what is the probability that the ID is correct? This is a question about the subset of eyewitnesses who identify a suspect. No other way of plotting the data (and no numerical summary of the data—not the point-biserial correlation coefficient nor any other statistic) provides the answer to that question more directly and more understandably than a visual plot relating suspect-ID accuracy to confidence. At a glance, it not only reveals how much suspect-ID accuracy changes as a function of confidence, it also shows how reliable eyewitness IDs are for each level of confidence. Therefore, we use this approach in our review of actual experiments, to which we now turn.

A Review and Reanalysis of Eyewitness Confidence and Accuracy Data

We begin with a reanalysis of some of the studies that were included in the Sporer et al. (1995) review, which used the correlation coefficient to quantify the confidence-accuracy relationship, and then we reanalyze subsequent data that were originally published as calibration curves or as receiver operating characteristic (ROC) curves.

A reanalysis of three of the original Sporer et al. (1995) meta-analysis experiments

What would the data that were reviewed by Sporer et al. (1995) suggest about the confidence-accuracy relationship if, instead of computing the point-biserial correlation for choosers, one simply plotted the data as a CAC plot? To find out, we contacted each of the authors of that article and requested the original data from the 30 experiments analyzed in their Table 1 so that the data could be replotted as CAC curves. Quite understandably, most of the data are no longer available. However, data from three of those 30 experiments, which are representative of the larger data set in terms of the point-biserial correlation, are still available and were kindly provided to us by J. Don Read. The three experiments are Experiments 1 and 2 from Read, Yuille, and Tollestrup (1992) and Experiment 3 from Read, Tollestrup, Hammersley, McFadzen, and Christensen (1990). The data from Experiment 2 of Read et al. (1992) were originally analyzed separately for two targets (a central suspect and a peripheral suspect)

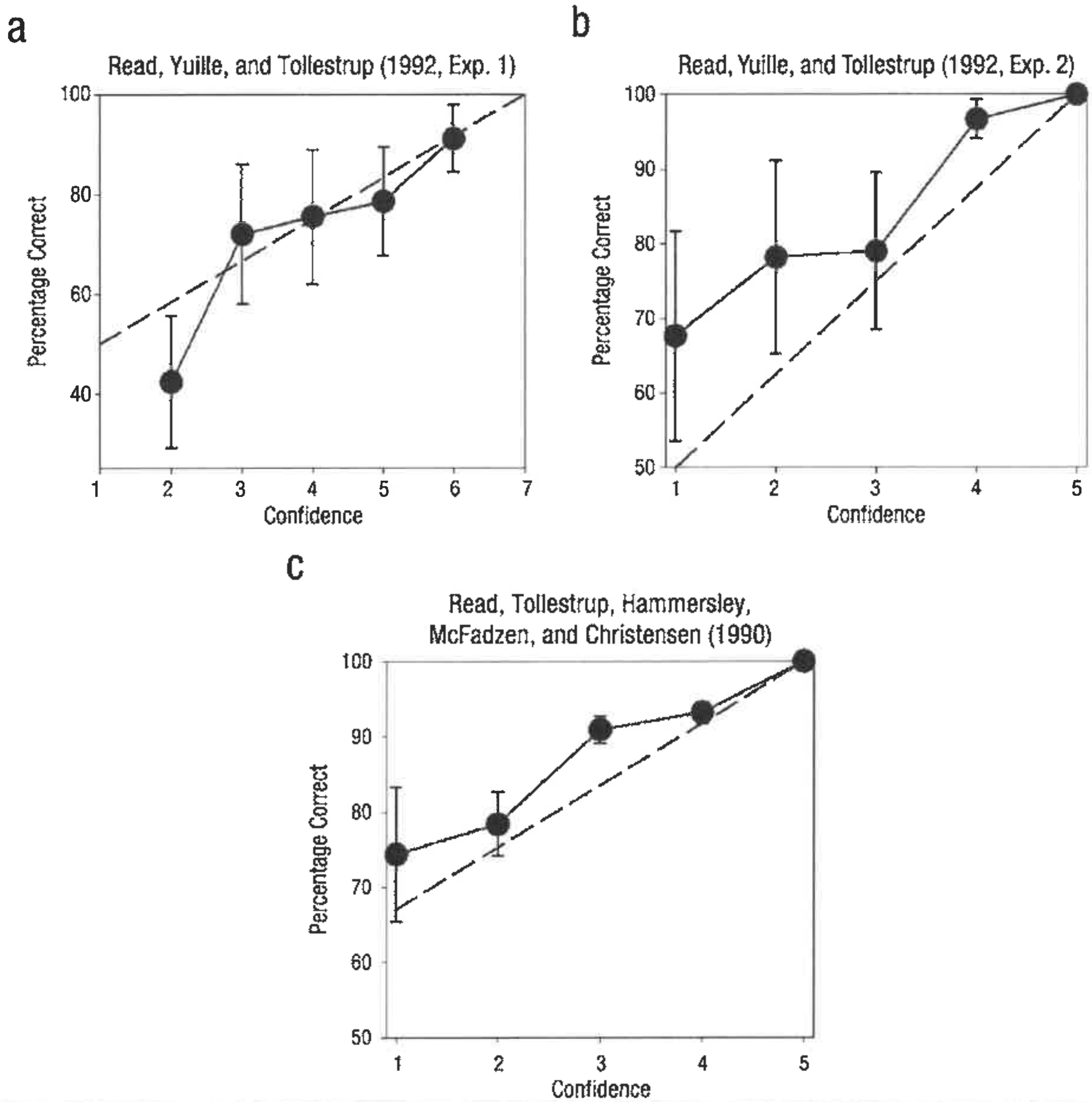


Fig. 3. Confidence-accuracy characteristic curves for three studies from the original Sporer et al. (1995) meta-analysis. The dashed lines indicate chance accuracy.

who were tested using different lineups, but we combined these data in our analysis because the results were quite variable when plotted separately because of the relatively small number of participants tested.

The point-biserial correlation coefficients for choosers in these three experiments were, respectively, .246, .511, and .359. The mean and standard deviation of these three values (M = .37, SD = .13) are similar to the corresponding values in the full data set (M = .41, SD = .16). In other words, these data appear to be reasonably representative of the larger set of studies analyzed by Sporer et al. (1995). What do these same data look like when plotted as a CAC curve? Figure 3 shows the CAC curves for the three obtainable data sets from the experiments analyzed in the original Sporer et al. (1995) meta-analysis. In each figure, a dashed diagonal line has been drawn for reference purposes. The line is drawn in such a way that the lowest confidence rating corresponds to chance suspect-ID accuracy and the highest confidence rating corresponds to 100% correct (perfect accuracy).

Note that chance accuracy corresponds to the level of performance that would be obtained if choosers randomly identified individuals from lineups. If an equal number of target-present and target-absent lineups were used, then chance accuracy for suspect IDs would be 50%. For example, if 1,000 "choosers" randomly sampled from six-person target-present lineups, about 1/6 of them would land on the guilty suspect. Thus, this group would, on average, identify (1 / 6) × 1,000 = 167 guilty suspects. Similarly, if another 1,000 choosers randomly sampled from six-person target-absent lineups, about 1/6 of them would land on the innocent suspect. Thus, this group would identify 1,000 fillers and an estimated (1 / 6) × 1,000 = 167 innocent suspects. In other words, each group would identify the same number of suspects. Half of the suspect IDs would be randomly made to guilty suspects and half would be randomly made to innocent suspects. Thus, suspect-ID accuracy would be 167 / (167 + 167) = .50. Generally speaking, when using CAC analysis, random chance accuracy is equal to the base rate of target-present lineups used in a study.

For the two experiments from Read et al. (1992) shown in Figure 3a and 3b, the base rate of target-present lineups was approximately 50%. Thus, random chance suspect-ID accuracy in these two studies was 50% correct (and, of course, perfect accuracy is 100% correct). For the experiment from Read et al. (1990) shown in Figure 3c, the base rate of target-present lineups (and, therefore, chance accuracy) was approximately 67%. In every plot we show (both here and for all subsequent calibration and ROC experiments we consider), the diagonal line represents the full range of performance from chance accuracy (usually 50%) to perfect accuracy (100%).

By any reckoning, the data shown in Figure 3 exhibit a strong relationship between confidence and accuracy.[1] This is true despite the fact that the very same data are associated with a mean point-biserial correlation between confidence and accuracy of .37, meaning that only 14% of the variance was accounted for (i.e., $.37^2 = .14$). The data in Figure 3a range from 42% correct for the lowest level of confidence (a rating of 2 on the 7-point scale) to 91% correct for the highest level of confidence (a rating of 6 on the 7-point scale). Participants in that particular study used confidence ratings of 1 and 7 too rarely to estimate performance associated with the lowest and highest possible confidence levels. A similar range of performance is evident for the data shown in Figure 3b (low-confidence accuracy = 62% correct; high-confidence accuracy = 100% correct). In Figure 3c, a smaller range is evident but only because low-confidence accuracy was fairly high because of the high target-present base rate in that study.

Note that the highest levels of confidence in Figure 3b and 3c are both associated with 100% accuracy. Averaged across the three experiments, accuracy associated with the lowest level of confidence was 61.4% correct (this score would be slightly lower had all three studies involved a 50% target-present base rate), whereas accuracy associated with the highest level of confidence was 97.0% correct. That is, according to these studies, which are associated with an average point-biserial correlation of only .37, low confidence implies low accuracy, whereas high confidence implies very high accuracy. These are some of the same studies that have helped to convince the legal system to disregard eyewitness confidence because the correlation between confidence and accuracy is low.

A reanalysis of later research using the calibration approach

Next, we review studies that reported calibration curves for choosers and were designed to create pristine testing conditions. Most of these studies did not report choosing rates for each filler ID in target-absent lineups, so we cannot be sure that a fair lineup was used in every case. Nevertheless, in these studies, the same decision rule was used to select the fillers and the replacement photo, so the results, considered in the aggregate, represent what would likely be observed when pristine testing conditions are used. Every one of these studies, many of which come from the Neil Brewer lab, has shown a strong relationship between confidence and accuracy, as the authors of these studies have repeatedly emphasized. The calibration plots in the published literature have generally counted only suspect IDs from target-present lineups while counting all filler IDs (not just estimated suspect IDs) from target-absent lineups. Here, we replot those same data in the form of CAC plots, which means we focused on correct suspect IDs from target-present lineups and (usually estimated) incorrect suspect IDs from target-absent lineups.

Recently, a number of studies using ROC analysis in the context of eyewitness identification have been published. These studies have used confidence ratings to construct the ROC, but most were not specifically concerned with the confidence-accuracy relationship. Still, these studies also provide the data needed to construct a CAC plot, so we included ROC studies as well. More specifically, we included in our analysis the calibration and ROC studies that met the following criteria:

1. The studies investigated recognition memory for faces (not recall of details).
2. The participants were adults.
3. The lineups were designed to be fair in that the replacement photo for target-absent lineups was chosen using the same decision rule that was used to choose the other fillers (we consider unfair lineups in a later section).
4. Confidence ratings were taken soon after the ID (5 minutes or less post-ID).
5. Memory was tested using a lineup.

Table 1. Confidence-Accuracy and ROC Studies Included in Our Review

Study	Experiment	Original figure or table	Notes
1*	Brewer, Keast, and Rishworth (2002)	Figure 1	
2*	Brewer and Wells (2006)	Figure 1	
3	Carlson and Carlson (2014)	Figure 1	Reanalysis of raw data that were supplied by first author; the 7-point confidence scale was reduced to a 3-point scale to reduce error variance; and we excluded a condition involving photos with an artificial feature (all faces in the lineup had a large black letter *N* sticker on one cheek) because it seemed far removed from the forensic situation.
4*	Carlson, Dias, Weatherford, and Carlson (2016)	Figure 2	
5*	Dobolyi and Dodson (2013)	Figure 2	
6*	Dodson and Dobolyi (2016)	Figure 1	Multiple numerical confidence scales were used; all were converted to 0%-to-100% scales.
7*	Horry, Palmer, and Brewer (2012)	Figure 3	
8	Juslin, Olsson, and Winman (1996)	Figure 4	This study used a 75% target-present base rate; accuracy scores were estimated as described in the appendix of Wixted, Read, and Lindsay (2016).
9*	Keast, Brewer, and Wells (2007)	Figure 1	This study reported data for adults only, which were a subset of the data in Brewer and Wells (2006).
10	Lindsay, Nilsen, and Read (2000)	Table 3	The 11-point confidence scale was reduced to a 3-point scale (*low*, *medium*, and *high*) to reduce error variance.
11*	Mickes (2015)	Figure 2; Figure 4	Data were collapsed across the recollection and no-recollection conditions of Experiment 1 because too few low-confidence IDs were obtained to yield stable accuracy estimates; simultaneous lineup data shown in Figure 4 (Experiment 2) are also included in our plot.
12*	Mickes, Flowe, and Wixted (2012), Experiment 1a and 1b combined	Figure 6a	We reanalyzed the raw data (ROC data were reported in the original article).
13*	Palmer, Brewer, Weber, and Nagesh (2013), Experiment 1	Figure 1	
14*	Palmer, Brewer, Weber, and Nagesh (2013), Experiment 2	Figure 3	
15	Read, Lindsay, and Nichols (1998), Experiment 3	Figure 6.4	Data were collapsed across the prewarned and nonwarned conditions to reduce random error; raw data provided by the first author were reanalyzed.
16*	Sauer, Brewer, and Wells (2008)	Table 3	
17*	Sauer, Brewer, Zweck, and Weber (2010)	Figure 1	
18*	Sauerland and Sporer (2009)	Figure 3	
19	Smith and Flowe (2014)	Figure 2	We reanalyzed the ROC data reported in the original article.
20*	Weber and Brewer (2004)	Figure 5	Mini-lineups (four members) were used in this experiment.

Note: The 15 studies marked with an asterisk all used a 100-point confidence scale.

Relevant studies were identified by searching the Web of Science database using the keywords "calibration," "confidence," and "eyewitness identification." In addition, we searched references cited by the identified studies, and we examined all studies that later cited the articles included in our review. We do not claim this to be an exhaustive review, but it is undoubtedly a large and representative sample of calibration studies. The studies that satisfied these criteria and that were included in our review are listed in Table 1. We also included Read, Lindsay, and Nichols (1998) and D. S. Lindsay, Nilsen, and Read (2000) even though those authors did not specifically present their data as a calibration curve or as an ROC curve. However, J. Don Read provided us with the raw data from Read et al. (1998), and D. S. Lindsay et al. (2000) presented their data in enough detail that a CAC plot could be constructed. We further included the adult sample from Keast, Brewer, and Wells (2007) even though their study used a subsample of adult participants who were tested by Brewer and Wells (2006), which is also included in our review.

Panels (a) through (s) of Figure 4 present the CAC plots from the calibration studies and ROC studies that we identified. Some of the studies included their raw data, making it possible to directly compute suspect-ID accuracy. For studies that did not, we precisely estimated accuracy scores from their calibration plots using WebPlotDigitizer (http://arohatgi.info/WebPlotDigitizer/) and converted the reported accuracy score that included filler IDs to one that included only suspect IDs. This was accomplished by taking the reported accuracy score for a given level of confidence, $a1$; converting it to an odds score, o, where $o = a1 / (1 - a1)$; and then computing suspect-ID accuracy, 100% × $a2$, using the formula $a2 = o / (o + 1 / n)$, where n = lineup size. An example showing how this works is presented in Appendix B. Figure 4 does not show error bars because it was not possible to compute them when the data were estimated. However, an aggregate plot presented later in Figure 5 provides an indication of the consistency across studies. Note that most of the studies on the confidence-accuracy relationship that reported only the point-biserial correlation could not be included in our review because there is no way to produce a CAC plot when all that is known is a correlation coefficient. Overall, four studies that originally reported a point-biserial correlation coefficient were considered here: the three studies shown in Figure 3 and the study by Read et al. (1998) shown in Figure 4n. As noted above, it was possible to include these studies because J. Don Read still had (and provided us with) the raw data.

Most of the studies we review reported data from multiple conditions, so for each study shown in Figure 4, we have plotted the results from the individual conditions on the left and the results aggregated across conditions on the right. The results are presented alphabetically by first author, except for the D. S. Lindsay et al. (2000) and Sauerland and Sporer (2009) studies, which are both shown in the final panel (Fig. 4s) because they had only one condition each. Figure 4b is based on the same data we used earlier to illustrate calibration curves for choosers (Fig. 2a). Generally speaking, the average plots on the right for the studies with multiple conditions in Figure 4 are representative of the individual-condition plots on the left, so the bottom-line story from those studies can be most easily appreciated by scanning the plots on the right. It is visually apparent that in most cases, high-confidence accuracy is very high (95%–100% correct), whereas low-confidence accuracy is obviously lower.

Fifteen of the relevant studies reported their data on a 100-point confidence scale. Most reported their results using the following scale: 0–20, 30–40, 50–60, 70–80, and 90–100. In a few studies, a 6-point scale was used consisting of 0%, 20%, 40%, 60%, 80% and 100% confidence. For those, we collapsed the 0% and 20% ratings together to create a 5-point scale so the data could be averaged with data from the other studies using a 100-point scale. Across those 15 experiments, the average accuracy of a low-confidence (0–20) ID was 63.7% correct (range = 37.5%–83.3%), whereas the average accuracy of a high-confidence (90–100) ID was 97.1% correct (range = 94.2%–99.7%). The resulting aggregate CAC curve is shown in Figure 5a.

Overall, the data from the calibration studies reviewed here tell the same story as the data from the experiments included in the original Sporer et al. (1995) meta-analysis that we were able to reanalyze (Fig. 3). Confidence is highly predictive of accuracy in the straightforward sense that low-confidence suspect IDs are error prone (though often well above 50% chance, so such IDs are somewhat probative of guilt) whereas high-confidence suspect IDs are largely, but not perfectly, accurate. Moreover, these data indicate that, with respect to suspect-ID accuracy, eyewitnesses are, if anything, *under*confident (not overconfident). Note that all of these studies were methodologically similar to the earlier studies that were reviewed by Sporer et al. (1995), which are the studies that have helped to convince the legal system to increasingly disregard eyewitness confidence. What differs is how the data are analyzed, and that difference changes the story of the relationship between eyewitness confidence and accuracy as it is currently understood by the legal system (based largely on the point-biserial approach).

Figure 5b shows the average calibration plot (counting filler IDs from target-present lineups as errors). Although the data shown in Figure 5a are of most interest to judges and juries, the data shown in Figure 5b are certainly of interest to scientists. This plot is relevant to the question of how well eyewitnesses can express confidence in a way that corresponds to their subjective impression of accuracy. Any viable theory of eyewitness confidence would have to accommodate these data as well. Remarkably, the data exhibit almost perfect calibration (cf. Juslin et al., 1996).

Unfair lineups

As indicated earlier, our conclusions about the relationship between confidence and accuracy apply to initial IDs made from fair lineups without undue influence from a lineup administrator. A fair lineup is one in which everyone in the lineup resembles the perpetrator to the same approximate degree, so the suspect would not be identified more often than chance by a group of mock witnesses provided with the perpetrator's description.

The situation is undoubtedly different when unfair lineups are used. An unfair lineup is one in which the suspect stands out from the fillers such that the suspect

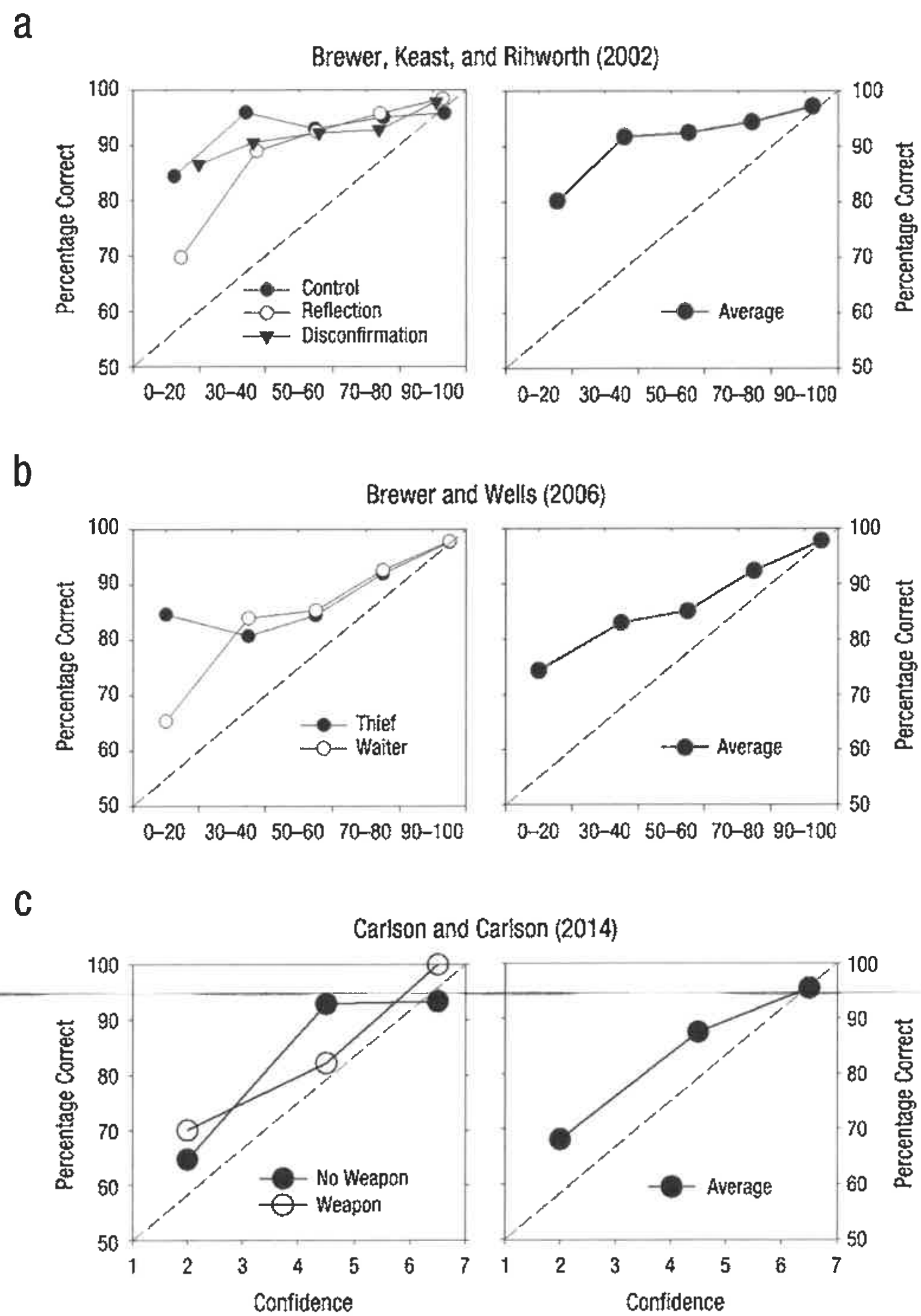


Fig. 4. (continued)

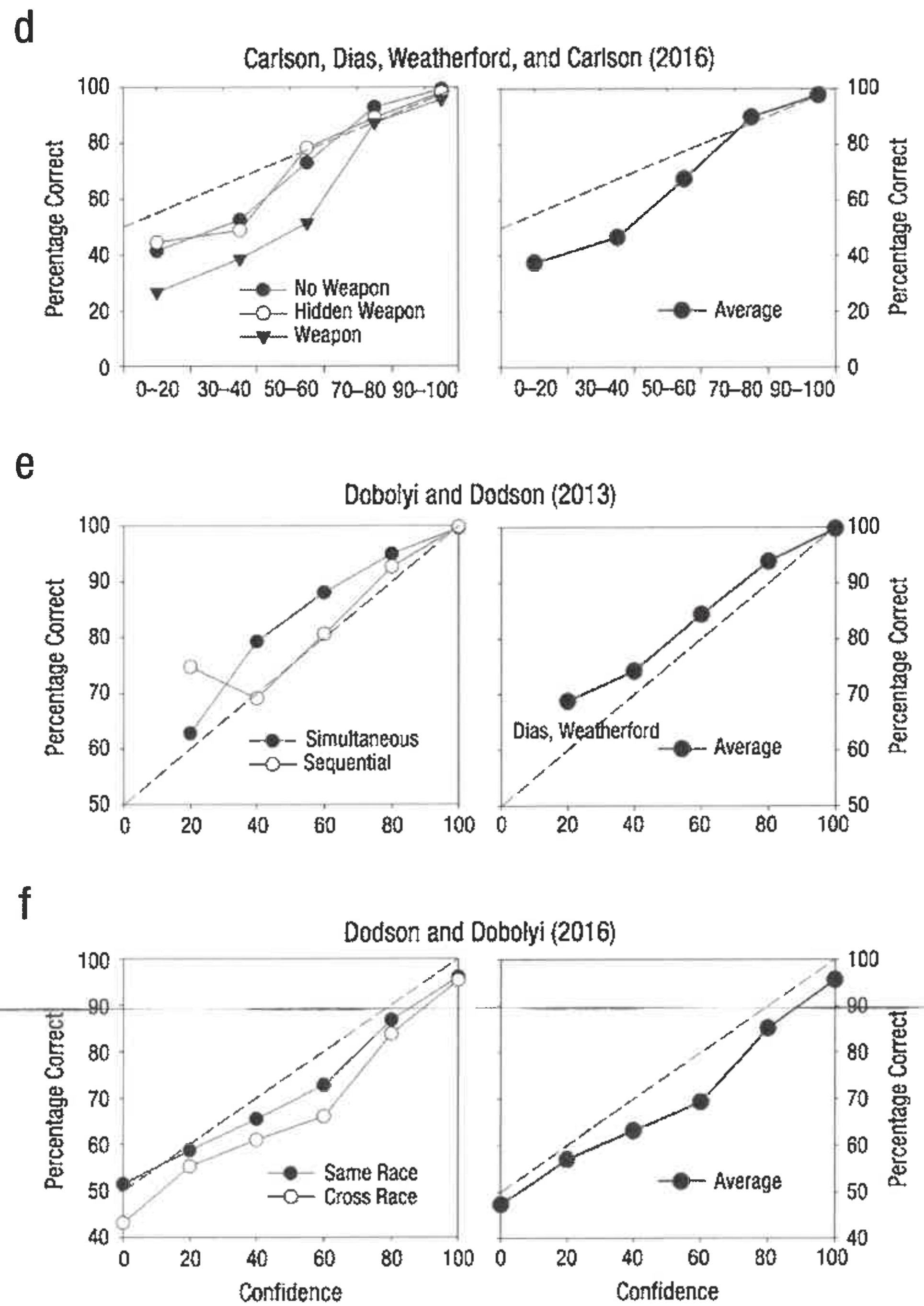


Fig. 4. (continued)

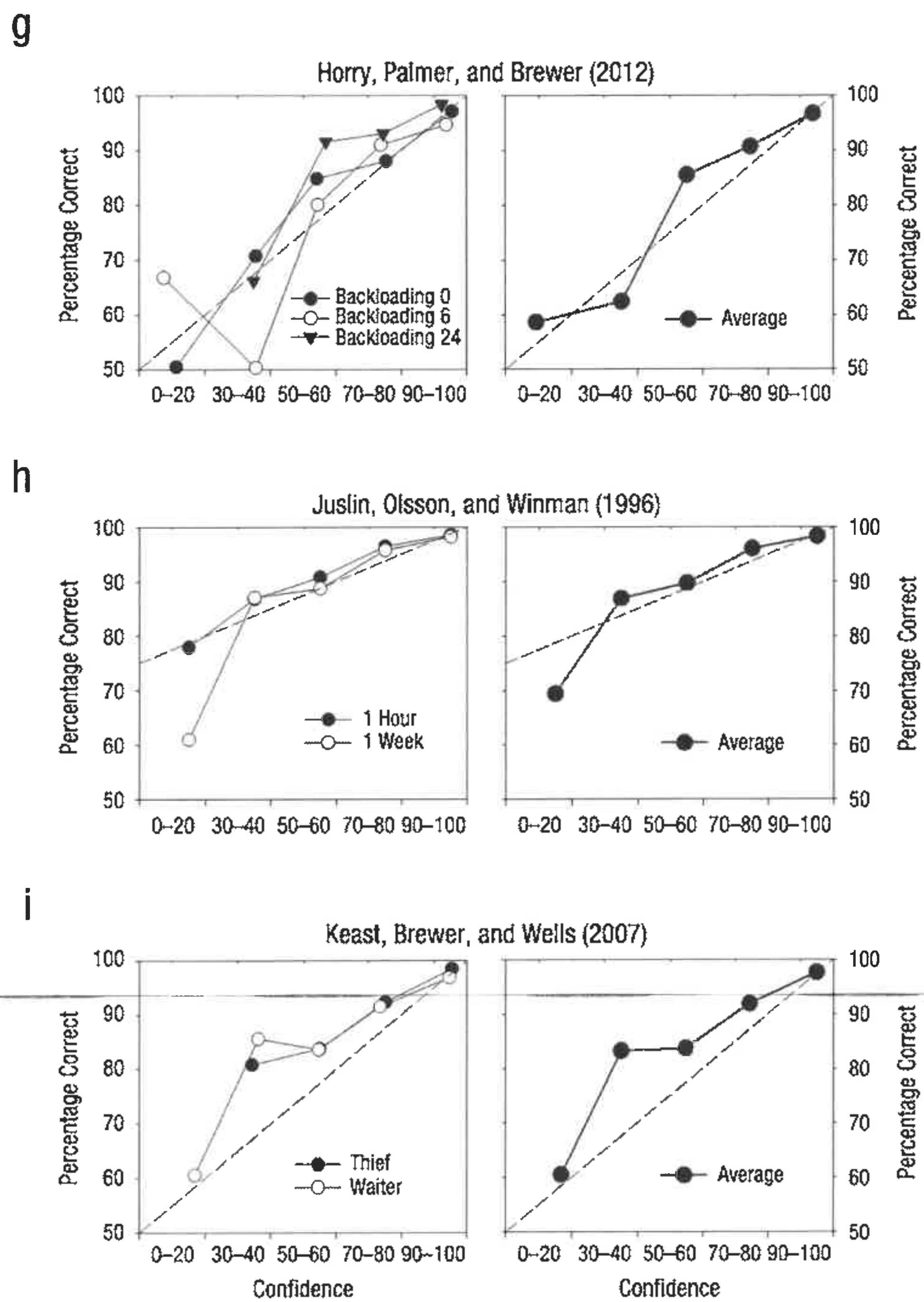


Fig. 4. (continued)

Plaintiff's Trial Exhibit 290
Page 24 of 56

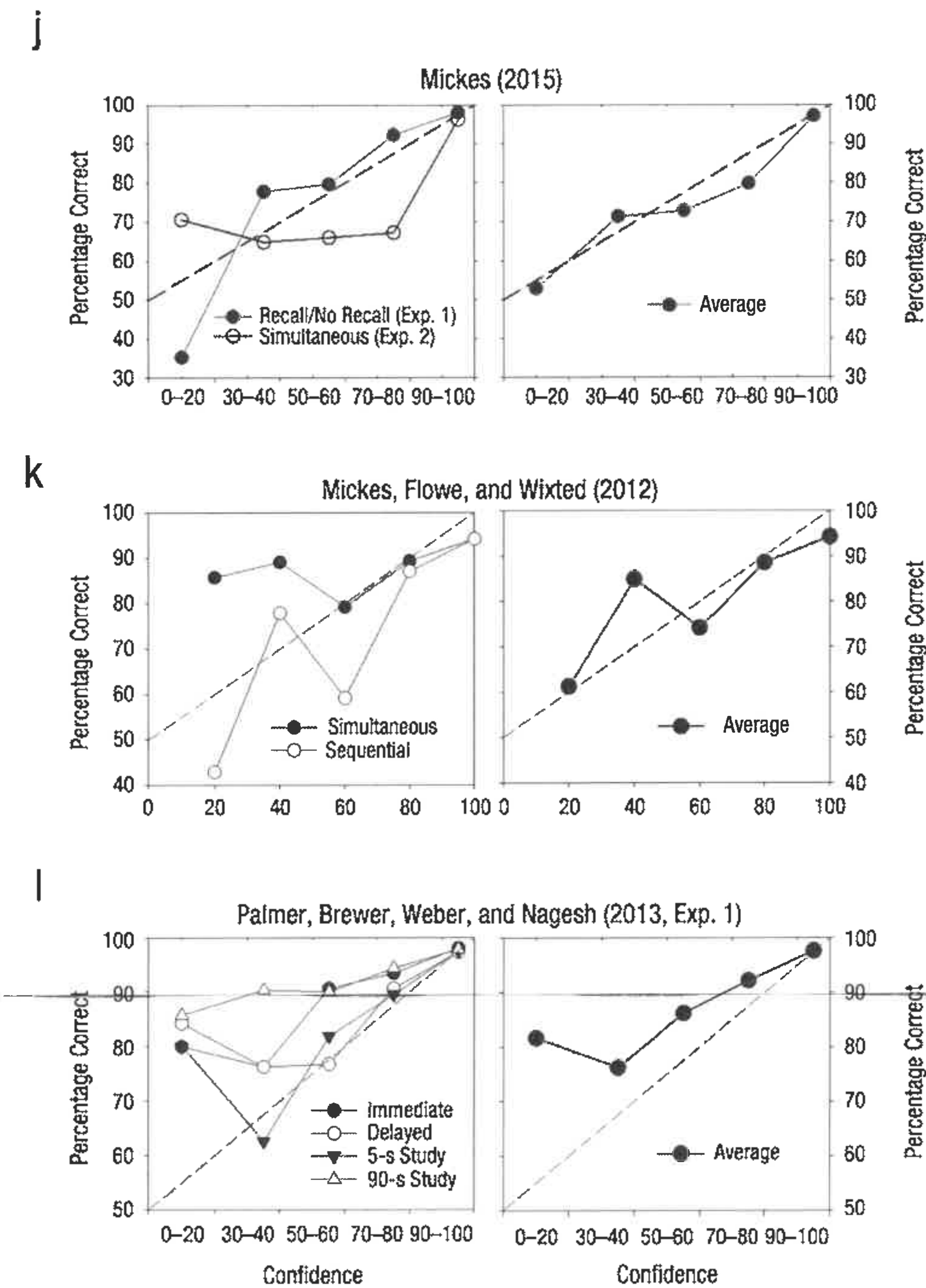


Fig. 4. (continued)

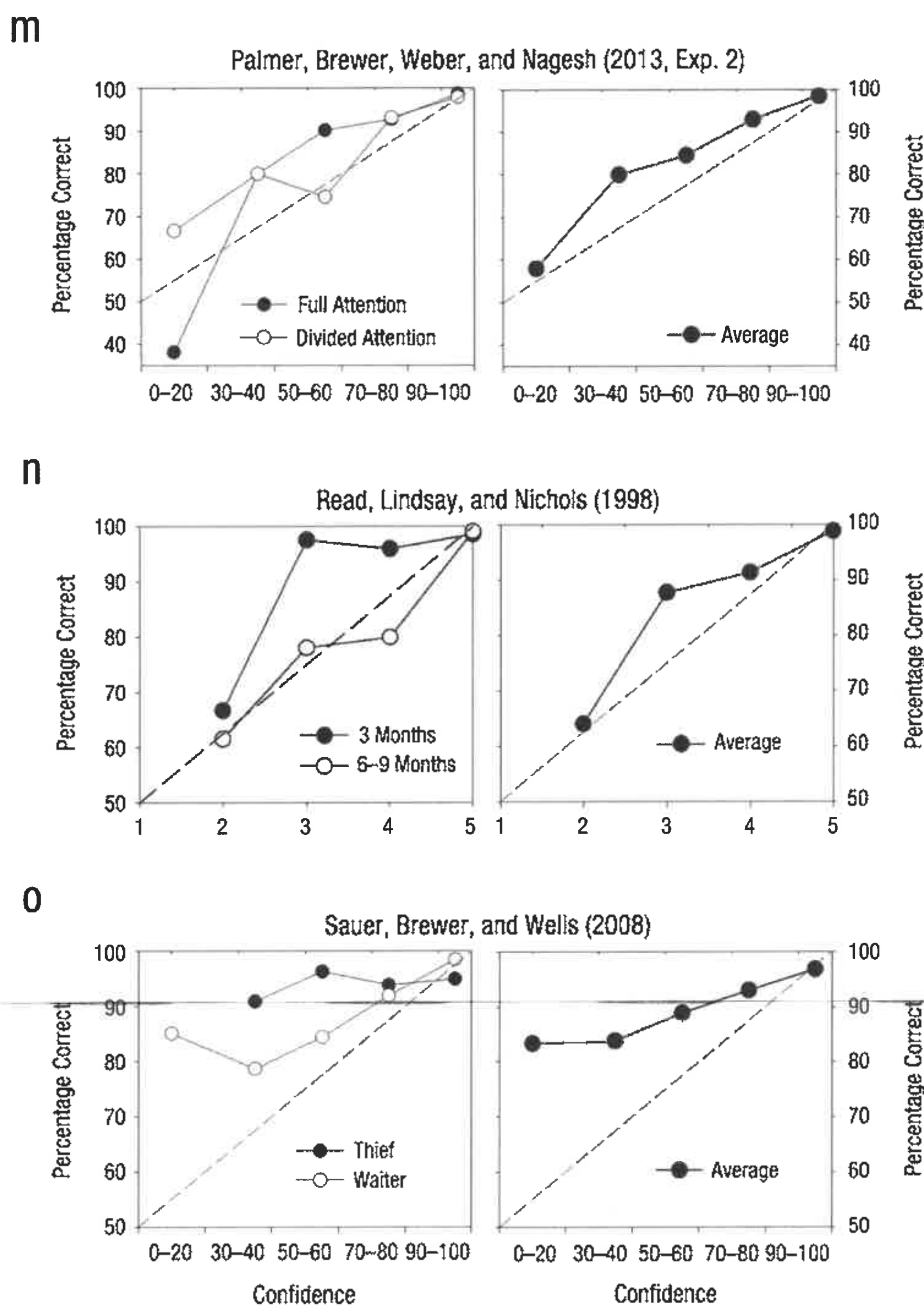


Fig. 4. (continued)

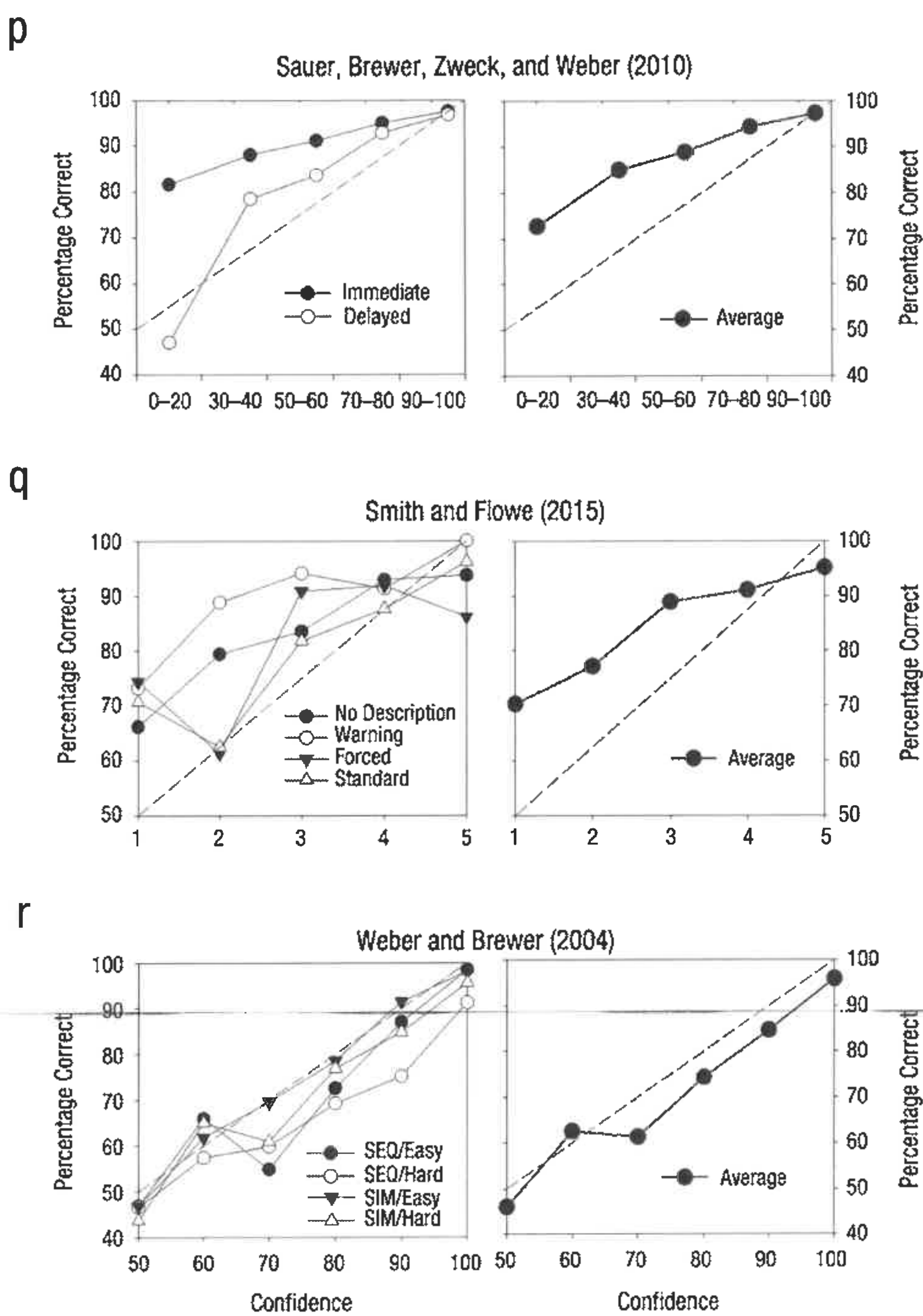


Fig. 4. (continued)

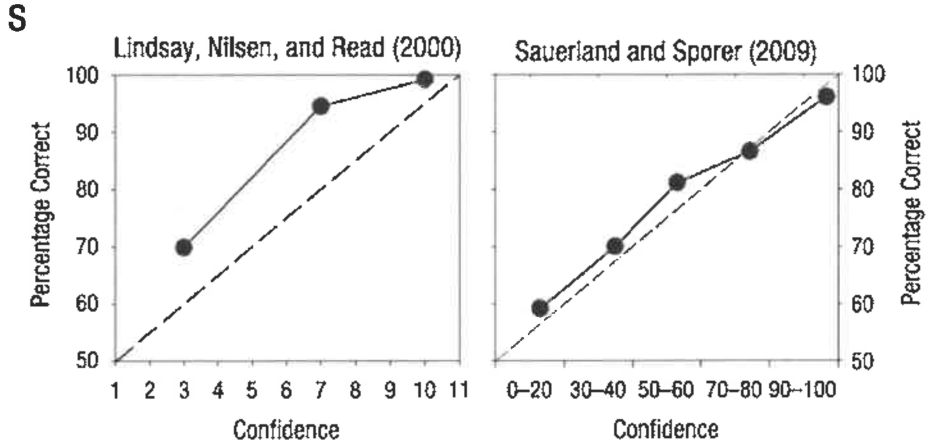


Fig. 4. Suspect-ID accuracy (percentage correct) as a function of confidence for the 20 studies listed in Table 1. SEQ = sequential; SIM = simultaneous.

(innocent or guilty) resembles the perpetrator to a noticeably greater extent than the fillers. It is well known that an unfair lineup leads to a higher rate of suspect identification and higher confidence in that identification, whether or not the suspect is the perpetrator (Fitzgerald, Price, Oriet, & Charman, 2013; R. C. L. Lindsay & Wells, 1980; Wells, Rydell, & Seelau, 1993). It stands to reason that unfair lineups would also reduce the utility of eyewitness confidence and would decrease the reliability of high-confidence IDs. In the extreme, placing the perpetrator's identical twin in a target-absent lineup would undoubtedly yield many incorrect high-confidence IDs of the innocent suspect, wreaking havoc on the accuracy of high-confidence suspect IDs.

Gronlund, Carlson, Dailey, and Goodsell (2009) conducted a large-scale investigation into the diagnostic accuracy of simultaneous and sequential lineups using target-absent lineups in which the designated innocent suspect resembled the perpetrator more than the fillers did. They also varied how much the picture of the guilty suspect resembled what the perpetrator looked like while committing the crime. In some conditions of that experiment, performance was near chance (e.g., when the perpetrator's appearance had substantially changed and the

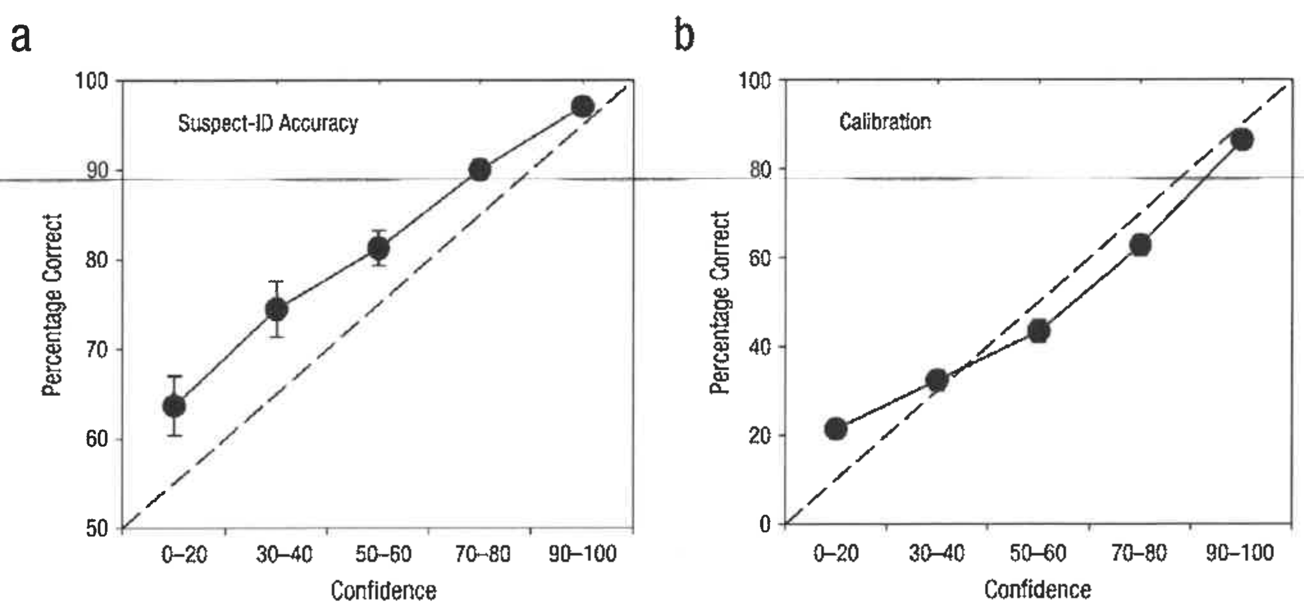


Fig. 5. Suspect-ID accuracy averaged across 15 studies with comparable scaling on the confidence (x-) axis (a), and the same data plotted as a calibration curve (b). The studies included in this analysis, which all used a 100-point confidence scale, are indicated in Table 1 with an asterisk.

Plaintiff's Trial Exhibit 290
Page 28 of 56

innocent suspect looked a lot like the original view of the perpetrator). In other conditions, performance was above floor, and Gronlund et al. (2012) reported confidence-based ROC data for those conditions. Collapsed across simultaneous and sequential lineups, the target-absent lineups from the conditions they analyzed were still unfair in the sense that the innocent suspect more closely resembled the perpetrator than the fillers did (so the innocent suspect was identified with much higher probability than the individual fillers were). Thus, these data can be used to gain some insight into the confidence-accuracy relationship when unfair lineups are used. Figure 6a presents the CAC plots for simultaneous and sequential lineups from Gronlund et al. (2012). The data exhibit a strong relationship between confidence and accuracy, but high-confidence accuracy (88% correct) is noticeably lower than it has been for the fair lineups considered to this point. The lower accuracy score for high-confidence IDs presumably reflects the impact of lineup unfairness. Note that the other conditions in their experiment, which yielded chance performance (because in some cases the photo of the innocent suspect resembled the perpetrator more than the photo of the guilty suspect did) would clearly wreak havoc on the confidence-accuracy relationship.

A clear illustration of the effect of unfair lineups can be observed by analyzing some of the data reported by Mickes, Flowe, and Wixted (2012). In their Experiment 2, the innocent suspect in the target-absent lineups was an altered photo of the perpetrator himself. The perpetrator's photo was altered using Photoshop to change the hair color, skin tone, nose shape, and face shape. Because these changes were all relatively minor, this experiment approximated a situation in which target-absent lineups contained a near twin of the perpetrator. As might be expected, the researchers' ROC analyses indicated that overall performance was rather poor. For simultaneous lineups, the overall correct-ID rate was .50 and the false-ID rate was .26 ($d' = 0.63$). For sequential lineups, the correct-ID rate was .42 and the false-ID rate was .22 ($d' = 0.55$). Of more interest for present purposes are the CAC plots shown in Figure 6b. Obviously, the relationship between confidence and accuracy is weaker than what is observed for fair lineups. Perhaps even more importantly, high-confidence suspect-ID accuracy is quite low (near 70% correct for simultaneous lineups and sequential lineups). As noted earlier, Sučić et al. (2015) also arranged unfair lineups, and the data from that study (expressed as a CAC plot) are shown in Figure 6c. Once again, high-confidence accuracy (only 85% correct) falls well below what is typically observed when fair lineups are used.

Two recent studies are particularly informative because they directly compared fair versus unfair lineups. Wetmore et al. (2015) tested participants using six-person simultaneous lineups either immediately after watching a mock-crime video or following a 48-hour delay. In their conditions in which the innocent suspect had only moderate similarity to the perpetrator (what they referred to as the "InnocentWeak" condition), some lineups were fair and others were biased against the innocent suspect. Their Table 2 presented choosing rates for the designated innocent suspect in each condition (fair vs. biased), so it was possible to determine that their manipulation of lineup fairness was successful. That is, the innocent suspect was disproportionately chosen over the other fillers in the biased condition only. Figure 6d presents the CAC results from that study collapsed across the retention-interval manipulation. As would be expected, the data from the fair condition are similar to the data presented earlier in Figure 4. Specifically, confidence is a strong predictor of accuracy, and high-confidence accuracy is very high (100% correct for confidence ratings of 7; 96% correct for confidence ratings of 6). However, in the biased condition, high-confidence accuracy is far lower (80% correct for confidence ratings of 7; 75% correct for confidence ratings of 6).

A similar pattern was evident in a recent study by Colloff, Wade, and Strange (2016). Participants watched a video of a perpetrator who had a distinctive feature, such as a black eye. In the unfair condition, the distinctive feature appeared only on the suspect in both target-present and target-absent lineups, not on any filler. Thus, whether innocent or guilty, the suspect stood out. In their fair conditions, by contrast, the distinctive feature either was present on all lineup members or was covered up for all lineup members (the data from several conditions in which the distinctive feature was added to or eliminated from all lineup members were very similar and have been averaged together here). As is apparent in Figure 6e, for the unfair condition, high-confidence accuracy was very low (~66% correct) and was much lower than high-confidence accuracy in the fair conditions (~86% correct). Although the effect of lineup fairness on high-confidence accuracy was consistent with other findings, high-confidence accuracy in the fair condition was noticeably lower than the ~95% correct levels of accuracy typically observed in the other studies reviewed here (e.g., Fig. 5a). Because there is no obvious reason for the observed difference, this result serves as a reminder that the determinants of high-confidence accuracy are not fully understood and that more research is needed to identify the conditions under which high-confidence accuracy can be compromised even when fair lineups are used.

These findings underscore the critical point that our claims about the relationship between confidence and accuracy (and, in particular, the very high level of accuracy usually associated with high-confidence suspect IDs) apply to fair lineups, not to unfair lineups. As noted by

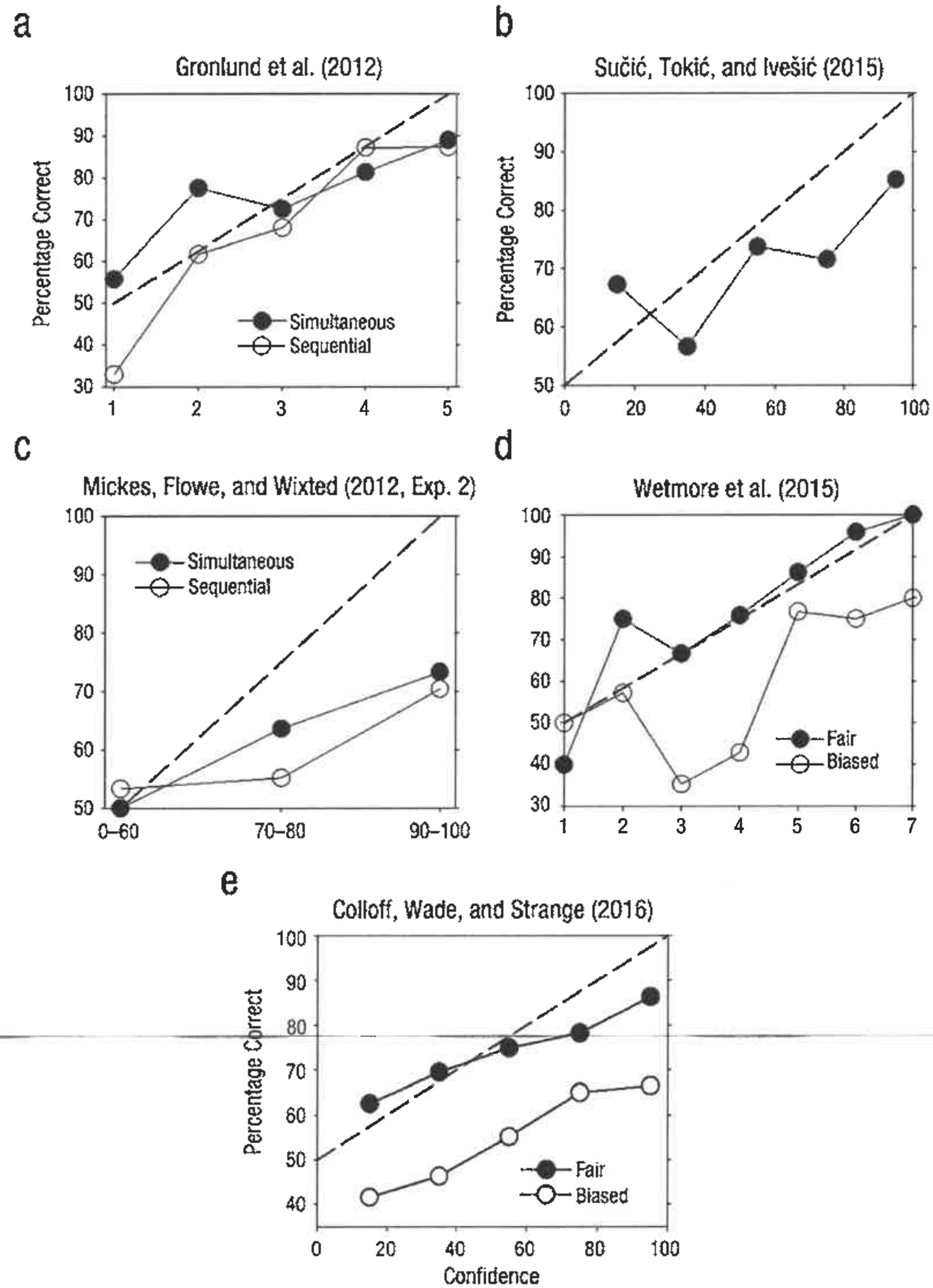


Fig. 6. Confidence-accuracy characteristic plots from studies that used unfair lineups.

Brewer and Palmer (2010), other circumstances in which the confidence-accuracy relationship may be degraded include (a) when the eyewitnesses are children (e.g., age 13 or younger), (b) when confidence ratings are not taken contemporaneously with the ID but are instead retrospective, and (c) when witnesses reject the lineup.

Plaintiff's Trial Exhibit 290
Page 30 of 56

Police Department Field Studies

The advantage of a mock-crime study such as the ones considered above is that the experimenter knows if a suspect ID is correct or incorrect, thereby allowing a direct computation of suspect-ID accuracy. In a police department field study, by contrast, it is not known if a suspect ID is correct or incorrect. Thus, although one can measure how often high-confidence and low-confidence IDs are made to suspects and fillers, a direct calculation of suspect-ID accuracy as a function of confidence is not possible. Nevertheless, indirect information about suspect-ID accuracy as a function of confidence can be obtained if (a) the perpetrator is a stranger to the witness (so the suspect in the lineup is not chosen because of preexisting familiarity), (b) the lineup is fair (so the suspect is not chosen because he or she stands out), and (c) blind administration is used (so the suspect is not chosen by the witness because of administrator influence). Under those conditions, the only way that the witness can land on the suspect with a probability that exceeds $1/n$, where n is lineup size, is if the suspect matches the memory of the witness. Except in rare cases in which an innocent suspect bears an uncanny resemblance to the perpetrator despite the fact that the lineup procedure was pristine, a strong memory-match signal would usually happen because the suspect actually is the perpetrator. Thus, if confidence is predictive of accuracy in the real world, suspect IDs should occur with probability greater than $1/n$, and that probability should increase as a function of confidence. To our knowledge, only two police department field studies have used fair lineups that were blindly administered and also reported confidence data. Both of these studies yielded data suggesting that high-confidence IDs are highly reliable, whereas low-confidence IDs are much less reliable (just as the lab data summarized in Fig. 5a would suggest).

Hennepin County police department field study

Klobuchar, Steblay, and Caligiuri (2006) conducted a pilot study of 206 actual eyewitnesses who were tested using six-person sequential photo lineups in four municipal police departments in Hennepin County, Minnesota. The lineups were not specifically tested for fairness but were presumably fair because department policy required the use of photographs depicting individuals of similar age, skin color, complexion, hairstyle, and build. The lineups were administered by an officer who was blind to the suspect's identity, and confidence was recorded in the witness's own words. Some lineups contained a suspect previously known to the witness, whereas other lineups contained a suspect previously unknown to the witness. The key measure was the frequency of ***jump-out IDs***, which are rapid IDs accompanied by expressions of absolute certainty. In other words, jump-out IDs are high-confidence IDs.

Of 175 choosers in this study, 96 (55%) made jump-out IDs. Remarkably, 99% of these IDs were made to suspects, not fillers, which is to say that only one of the 96 jump-out IDs was made to a filler. From their Table 5, it was possible to determine that 26 of the jump-out IDs were made to strangers, and 70 were made to suspects previously known to the eyewitness. The stranger data are of interest here. The one jump-out ID that landed on a filler occurred in a stranger lineup (Nancy Steblay, personal communication, April 25, 2016); thus, 25 out of 26 jump-out IDs in stranger lineups (96%) landed on the suspect.

Keep in mind that there were 5 times as many fillers as suspects in any given lineup, so random responding for jump-out IDs would result in $26 \times (5 / 6) \approx 22$ filler IDs (yet only one was actually observed) and only about $26 \times (1 / 6) \approx 4$ suspect IDs (yet 25 were actually observed). Thus, the number of suspect IDs made with high confidence in this study was far greater than would be expected by chance. It is possible that the lineups in the Hennepin County study were not fair lineups. But if they were fair lineups (as they were designed to be), it is hard to come up with a logical explanation for these results without assuming that high-confidence accuracy was close to perfect. IDs made with lower confidence in that study (non-jump-out IDs) landed on the suspect much less often, approximately 60% of the time. That is still much more often than would be expected by chance alone, so even these more error-prone suspect IDs appear to be somewhat probative of guilt. These results suggest a strong confidence-accuracy relationship that is not appreciably different from that revealed by the lab results depicted in Figure 5a.

Houston Police Department field study

Another recent police department field study was specifically designed, in part, to examine the information value of eyewitness confidence (Wixted et al., 2016). In this study, eyewitness decisions were recorded from six-person photo lineups administered as part of criminal investigations in the Robbery Division of the Houston Police Department between January 22 and December 5, 2013. This study involved the administration of 348 simultaneous and sequential lineups, the investigators were unaware of the identity of the suspect in each lineup (i.e., double-blind administration was used), and the lineups involved suspects who were unknown to the eyewitnesses prior to the crime. Lineup fairness was examined for a random sample of 30 photo lineups by providing

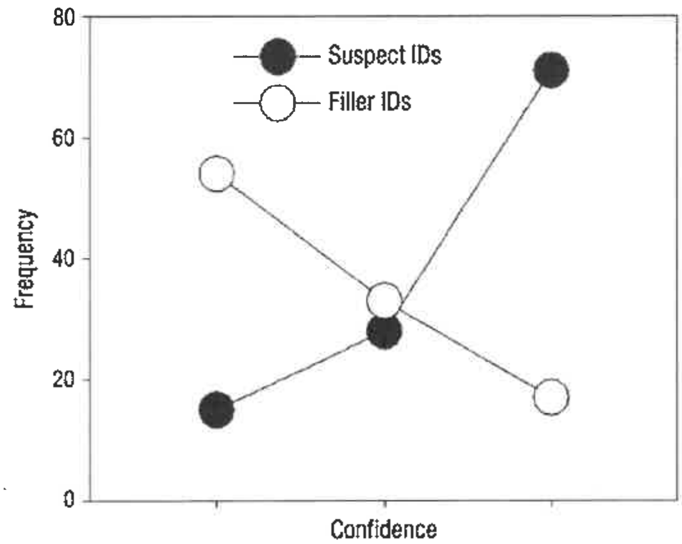


Fig. 7. Suspect IDs and filler IDs made with low, medium, and high confidence in the Houston Police Department field study (Wixted, Mickes, Clark, Dunn, & Wells, 2016).

the selected photo lineups to 49 mock witnesses and asking them to try to identify the suspect based only on the suspect's physical description. As noted above, in a fair six-person lineup, the suspect should be identified by a mock witness only 1/6 (.17) of the time. The mean proportion of suspect IDs made by the mock witnesses (.18) did not differ significantly from the expected value for a fair six-member lineup, $t(29) = 0.76$. Thus, according to this measure, the 30 lineups that were randomly selected were, on average, fair. For purposes of our analyses, we assumed that the remaining lineups were also fair. Eyewitnesses who made a suspect ID or a filler ID were asked to supply a confidence rating on a 3-point scale (*positive*, *strong tentative*, or *weak tentative*).

The critical results are reproduced in Figure 7. Obviously, most suspect IDs were made with high confidence, whereas most filler IDs were made with low confidence. This pattern again immediately suggests a strong confidence-accuracy relationship. Moreover, as with Klobuchar et al. (2006) and in agreement with lab studies (Fig. 5a), high-confidence IDs appear to have been highly accurate. Even though there were 5 times as many fillers as suspects in the police lineups used in this study, high-confidence IDs landed on the suspect 72 times and landed on a filler 17 times. Using perfectly fair lineups, one would expect 5 times as many high-confidence filler IDs as high-confidence suspect IDs. Thus, as a crude approximation, the 17 high-confidence filler IDs translate to an estimated $17 / 5 \approx 3$ high-confidence innocent-suspect IDs. If three of the 72 high-confidence suspect IDs were made to innocent suspects, it means that 69 of the 72 suspect IDs made with high confidence (96%) were correct. A formal signal-detection model fit to these data estimated high-confidence suspect-ID accuracy to be approximately 97% correct, whereas low-confidence suspect-ID accuracy was estimated to be closer to 50% correct. Again, these results are not dramatically different from the lab results summarized in Figure 5a.

Base Rates of Target-Present Lineups in the Laboratory and in the Real World

In most of the lab studies that we have considered here, the base rate of target-present lineups was 50%. An issue in generalizing from the lab to the real world is that the base rate of target-present lineups is unknown, and it is quite likely that the base rate will vary from one police department to another, or even from one detective to another, as a function of how much evidence an investigator requires before placing a possible suspect in a lineup (Wells, 1993). In order to explore the effect of different base rates, we used the data from Wetmore et al.'s (2015) fair lineups. Figure 6d showed a CAC on these data based on a 50% base rate. We created Bayesian curves called *prior-by-posterior curves* that map the probability that an identification of the suspect was accurate (i.e., that the suspect is the perpetrator) across all possible values of the base rate from 0% (all lineups had an innocent suspect) to 100% (all lineups had a guilty suspect). See Appendix C for a short tutorial on this

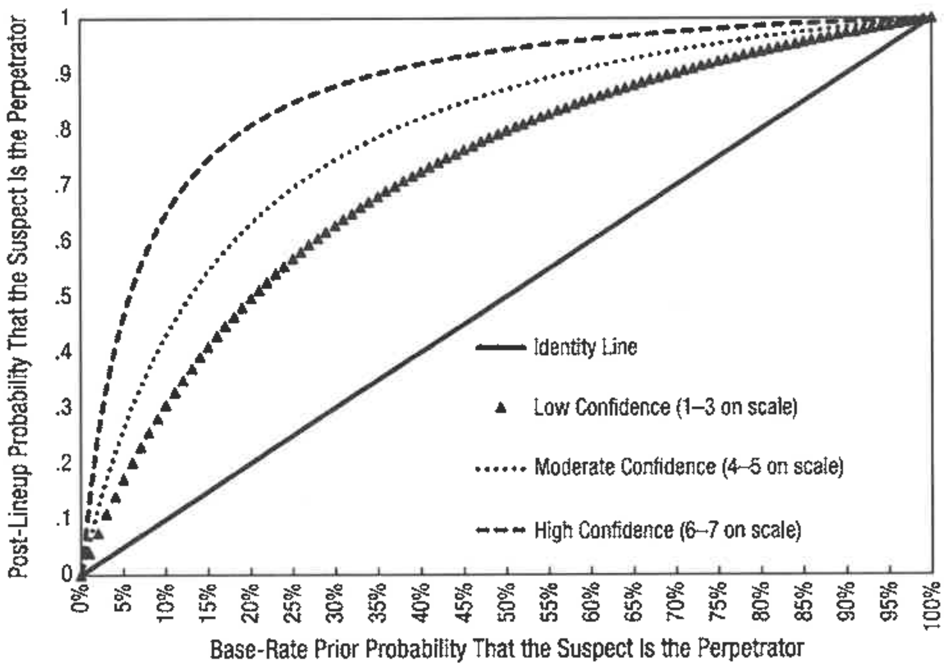


Fig. 8. Post-lineup probability that the suspect is the perpetrator as a function of the base rate of target present lineups and whether the ID was made with low, moderate, or high confidence. The data are from Wetmore et al.'s (2015) fair lineups.

Bayesian approach. With a sufficiently large sample size, we can create a curve for every level of confidence. But, because sample sizes can get small when every level of confidence is examined separately, we used three levels of confidence. The Wetmore et al. study used a 7-point confidence scale, so we collapsed confidence ratings of 1 through 3 into the category of low confidence, 4 and 5 into moderate confidence, and 6 and 7 into high confidence. These curves are shown in Figure 8.[2]

The solid line in Figure 8 is called an *identity line*, and it simply represents where the data would fall if the identification had no diagnostic utility. Clearly, all three curves are above the identity line and, as would be expected, the height of the curve for the high-confidence eyewitnesses is far above that of the curves for the moderate- and low-confidence witnesses.

Notice that the probability that the identified suspect is the perpetrator (which is the same as the probability that the witness is accurate) for high-confidence eyewitnesses remains relatively high (above 90%) until the base rate drops below 35%. Contrast that, however, with low-confidence witnesses, for whom the accuracy drops below 90% as soon as the base rate drops below 70%. In fact, whereas the high-confidence witnesses are still 90% accurate when the base rate is a mere 35%, the low-confidence witnesses drop all the way to a mere 63% accuracy if the base rate is 35%.

The data in Figure 8 underscore an important point made by Wells et al. (2015), namely that the base rate matters. Moreover, the base rate for lineups is a system variable. If a police department places a suspect who matches the perpetrator's description in a lineup on nothing more than a hunch, then the base rate of guilt in that jurisdiction is likely to be on the low side. Requiring at least some independent evidence of guilt (i.e., requiring more than just a hunch) will move a jurisdiction to the right on the base-rate dimension in Figure 8, thereby increasing the probability that an identification of a suspect is an accurate identification for all IDs made with any level of confidence.

What does law enforcement believe about the need to have evidence indicating that the suspect is likely to be the perpetrator before placing a suspect in the jeopardy of a lineup? A national survey indicated that more than one-third of U.S. crime investigators believed that they needed no evidence at all about the likely guilt of a person before placing that person in a lineup (Wise, Safer, & Maro, 2011). Behrman and Richards (2005) examined records from 306 lineups in Northern California in which a witness identified someone. They then coded how much evidence existed against the suspects before they were placed in a lineup. Behrman and Richards found that in 30% of the cases the evidence was "minimal," and in an additional 40% of the cases there was no pre-lineup

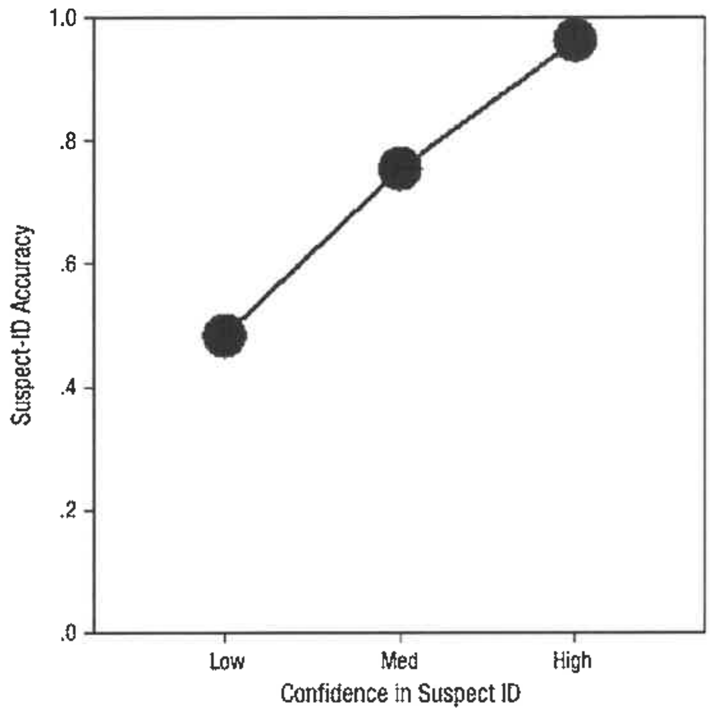


Fig. 9. Estimated suspect-ID accuracy as a function of confidence for the data from the Houston Police Department field study (Wixted Mickes, Clark, Dunn, & Wells, 2016). The estimates were based on a signal-detection model, which further estimated the target-present base rate of lineups to be 35%.

evidence at all. This does not tell us directly about what the base rates are in those jurisdictions, but it does not lend much confidence toward the idea that the base rate is high.

The base rate of guilt in lineups is generally assumed to be an unknowable variable in the real world. However, the signal-detection model used by Wixted et al. (2016) provided a principled estimate of the base rate in the Houston Police Department. The base-rate estimate that Wixted et al. reported—35%—is just that, an estimate, so it could be wrong. However, it is a principled estimate because it is based on a theory that has long guided thinking about recognition memory in other contexts. Moreover, it is a demonstration that, with the right theory, the base rate of guilt in a particular jurisdiction is not necessarily an unknowable value. The signal-detection model used by Wixted et al. (2016) may not be the right theory. And recall that the theory giving rise to this estimate assumed that the lineups were fair lineups based on an analysis of only a subset of the lineups. But the point is that base-rate information is not inherently unknowable, and the first principled estimate in a police jurisdiction came out surprisingly low. Fortunately, using pristine identification procedures, the laboratory data shown in Figure 8 suggest that, at a base rate of only 35%, confidence is highly predictive of suspect-ID accuracy, and high-confidence IDs are still quite accurate (about 90% in Fig. 8), whereas low-confidence IDs, despite having probative value, are highly error prone. Similarly high accuracy was obtained when the signal-detection model was used to estimate suspect-ID accuracy in the Houston field study assuming a 35% base rate of target-present lineups (Fig. 9). However, if the true base rate were much lower than that, then high-confidence IDs would begin to become highly error prone as well. Moreover, base rates likely differ from jurisdiction to jurisdiction, which means that some may fall well below the 35% estimate in Houston. Thus, conceptualizing the base rate as a system variable—and taking concrete steps to increase it—seems like a prudent strategy for law enforcement to consider.

Filler IDs and Non-IDs

In a police lineup, there are three possible decision outcomes: a suspect ID, a filler ID, and a non-ID (a rejection). To this point, we have focused on suspect IDs because those IDs are the ones that have often ended up putting an innocent person in prison, only to be exonerated by DNA evidence years later. Wells et al. (2015), however, pointed out that the other two decision outcomes—filler IDs and rejections—also provide useful information about the chances that the suspect is the perpetrator. Unlike identifications of the suspect, however, the information value of rejections and of filler identifications is exculpatory rather than incriminatory.

The fact that rejections provide exculpatory information is somewhat obvious and stems from the simple observation that witnesses are more likely to reject the lineup (make a non-ID decision) if it is a target-absent lineup than if it is a target-present lineup. But it is somewhat more difficult to intuit that filler IDs also provide exculpatory information. Empirically, it has long been recognized that filler identifications are more likely to occur in response to target-absent lineups than in response to target-present lineups (Wells & Lindsay, 1980). Accordingly, it makes sense that filler identifications would have exculpatory value. In effect, a witness who identifies a filler is offering an opinion that there is a filler in the lineup who looks more like the perpetrator than does the suspect. And, of course, that means that filler IDs are more likely to happen when the suspect is not the perpetrator than when the suspect is the perpetrator (Wells et al., 2015).

The information that rejections and filler identifications provide can be expressed at different levels of witness confidence using prior-by-posterior curves just as we did with identifications of the suspect. Prior-by-posterior curves for rejections are shown in Figure 10a, and the curves for filler identifications are shown in Figure 10b using the data from Wetmore et al. (2015). The dependent measure in Figures 10a and 10b is the probability that the lineup is a target-present lineup (i.e., that the suspect is the perpetrator). Notice that, unlike identifications of the suspect (see Fig. 8), both rejections and filler identifications produce curves that fall below rather than above the identity (no information) line. That is because both rejections and filler identifications have exculpatory information value rather than incriminatory information value.

In the case of rejections, which are shown in Figure 10a, the vertical axis is equivalent to the proportion of witnesses who made a correct decision to not identify anyone from the lineup. As can be seen, high-confidence rejections produce a curve that is farther below the identity line than the lines produced by moderate- or low-confidence rejections. This reinforces an important point, namely that lineup administrators should be obtaining confidence statements from witnesses for rejection decisions at the time of identification in addition to collecting confidence statements for identifications of suspects.

In the case of filler identifications, which are shown in Figure 10b, the vertical axis does not represent the proportion of witnesses who made a correct decision. After all, all filler identifications are errors. Nevertheless, filler identifications have information value because a filler identification is more likely to occur when the suspect is not the perpetrator (target-absent lineup) than when the suspect is the perpetrator (target-present lineup). Notice that the exculpatory value of filler identifications can be as high, and sometimes more so, than the exculpatory value of rejections. In other data sets (e.g., Brewer & Wells, 2006), high-confidence filler identifications were more exculpatory than were lower levels of confidence, whereas in the Wetmore et al. (2015) data, it was moderate-confidence filler identifications that were most informative in the exculpatory direction (with low- and high-confidence filler identifications being equally informative in the exculpatory direction). However, this might be due to the fact that there were very few high-confidence filler identifications in the Wetmore et al. data, making the high-confidence filler-identification curves somewhat unstable. Wells et al. (2015) argued that high-confidence filler identifications should generally be more exculpatory than lower-confidence filler identifications because high confidence filler identifications indicate stronger confidence by the witness that the filler is a better match to the perpetrator than is the suspect.

The fact that filler identifications have exculpatory value is an important observation in light of evidence that law enforcement agencies often fail to make records of filler identifications. In their analyses of police files to score the outcomes of photo lineups in actual cases, researchers have found that lineup administrators failed to make records of filler identifications but always made records of suspect identifications (Behrman & Davey, 2011; Tollestrup, Turtle, & Yuille, 1994). Consistent with this, in a recent national survey, U.S. law enforcement agencies admitted that they do not even prepare a report of a lineup if the witness does not ID the suspect (Police Executive Research Forum, 2013). In a controlled experiment, Rodriguez and Berry (2014) assigned research participants to the role of lineup administrators who were either blind to which lineup member was the suspect or knew which lineup member was the suspect and which were fillers. When participant-administrators were blind, they made records of all of the witnesses' identifications (both suspect IDs and filler IDs). When the participant-administrators were not blind, however, they commonly failed to make records of filler IDs. Hence, this is yet another argument in favor of why eyewitness identifications should be conducted using double-blind procedures. In the absence of double-blind procedures, the results can be selectively reported.

Another important point is that the exculpatory value of filler identifications and rejections (pointing toward innocence) is generally less than the incriminating value

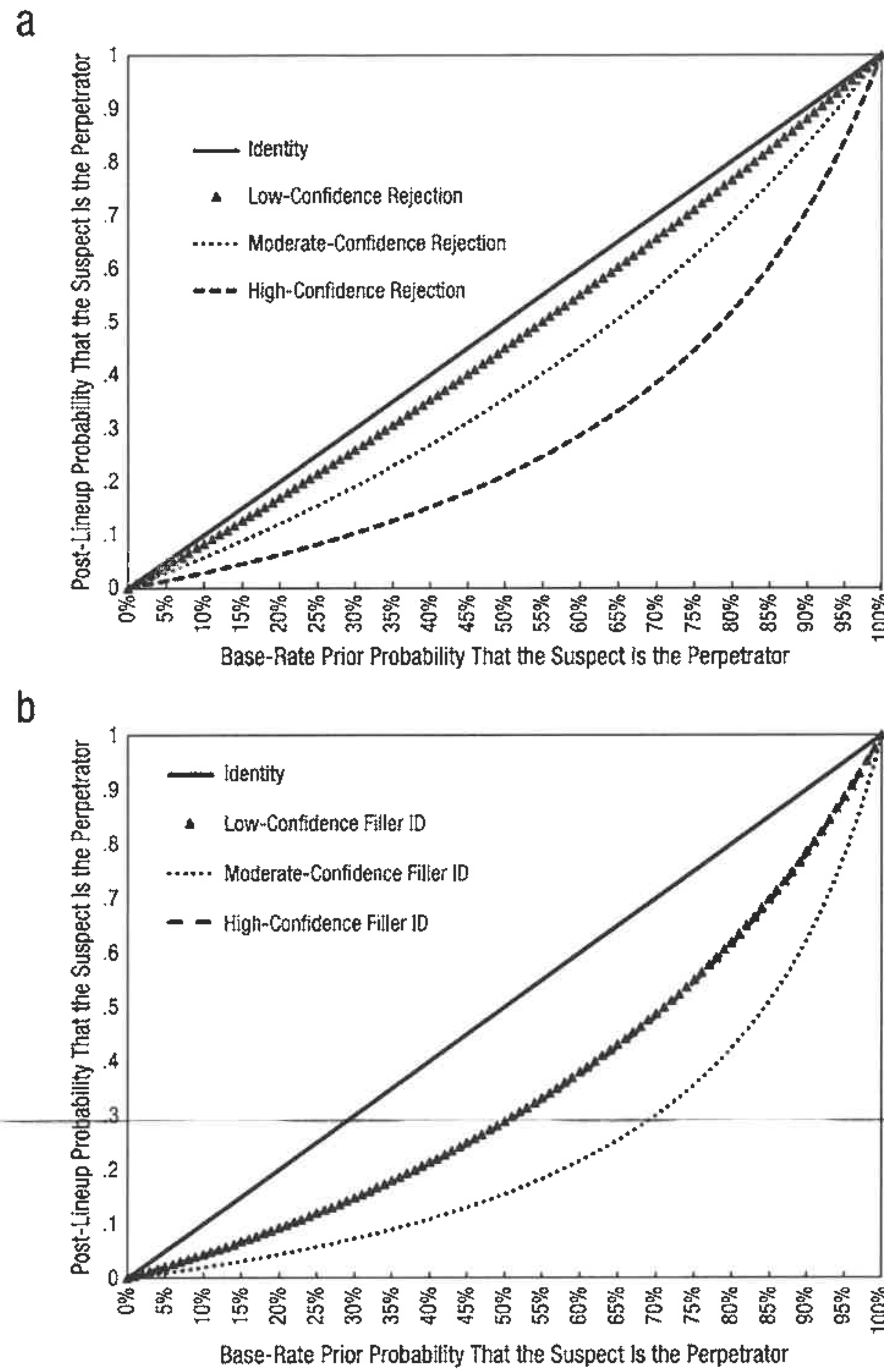


Fig. 10. Post-lineup probability that the suspect is the perpetrator as a function of the target-present base rate for lineups and the confidence of the witness who rejected the lineup (a) or identified a filler (b). The data are from Wetmore et al.'s (2015) fair lineups.

of identifications of the suspect (pointing toward guilt). This is apparent from noting that the area under the prior-by-posterior curves for identifications of the suspect (see Fig. 8) is greater than the area under the curve for rejections or identifications of fillers (Fig. 10). That type of pattern in the broader eyewitness-identification literature led Wells et al. (2015) to conclude that lineups are more effective for incriminating suspects than they are for exculpating suspects.

Theoretical Considerations

What explains the fact that under appropriate testing conditions, eyewitness confidence is such a reliable indicator of accuracy, but under other testing conditions it is not? We begin by discussing why confidence and accuracy ought to be related in the first place. Then, we discuss the ways in which non-pristine testing conditions manage to confound this relation.

The signal-detection-theory account of a strong accuracy-confidence relation

In a fair lineup administered in double-blind fashion, it will usually be the case that the only face in the lineup that will generate a strong memory-match signal is the face of the perpetrator (i.e., the face that created the memory trace in the first place). Except in rare cases of chance resemblance between an innocent lineup member and the perpetrator, no other face in the lineup should generate a strong memory-match signal because these other faces were not the source of the witness's memory. This is true whether the operative memory signal is the absolute strength of the match between the memory of the perpetrator and a single face in the lineup (without regard for the other faces in the lineup) or is instead the relative strength of that match compared to the match generated by the other faces in the lineup. Either way, only a guilty perpetrator is likely to generate a strong memory signal.

Presumably, witnesses have learned through the course of daily life that a strong memory signal is an indicator of high recognition accuracy (and therefore warrants a high-confidence ID), whereas a weak memory signal is an indicator of low recognition accuracy (and therefore warrants either a low-confidence ID or a lineup rejection). Thus, under pristine testing conditions, simply relying on the strength of the absolute or relative memory signal ought to result in a strong confidence-accuracy relation (Mickes, Hwe, Wais, & Wixted, 2011). These ideas can be formalized in terms of a simple signal-detection model (Fig. 11), which has long been used to conceptualize the strong confidence-accuracy relationship observed in list-memory tasks used by basic memory researchers. The model in Figure 11 is usually applied to word-list memory tasks, but the basic concepts also apply to decisions made from a lineup. A version of the model applied to lineups would be somewhat more complicated, but its basic predictions about the confidence-accuracy relationship would remain unchanged. Thus, for the sake of simplicity, we use the standard (list memory) version of the model to illustrate what it predicts about the confidence-accuracy relationship.

In the context of eyewitness memory, signal-detection theory specifies how face-memory strength is distributed across guilty suspects (targets) and innocent suspects and fillers (lures) in a fair lineup. As depicted in Figure 11, the mean and standard deviation of the target distribution are both greater than the corresponding values for the lure distribution (a common but not necessary assumption). The model assumes that a decision criterion is placed somewhere on the memory-strength axis, such that a positive identification is made if the memory strength of a face (target or lure) exceeds it. Each level of confidence is associated with its own decision criterion. The overall correct-ID rate is represented by the proportion of the target distribution that falls to the right of the leftmost decision criterion, and the overall false-ID rate is represented by the proportion of the lure distribution that falls to the right of the leftmost decision criterion. Our concern here is not with the overall correct- and false-ID rates but is instead with the frequency of confidence-specific correct and false IDs. As illustrated in Figure 11, high-confidence IDs occur when a face generates a strong memory signal, one that exceeds the rightmost decision criterion. For the specific example shown in that figure, high-confidence IDs will often occur for target faces (37% of target-present trials result in a correct high-confidence ID) but will rarely occur for non-target faces (only 2% of target-absent trials result in an incorrect high-confidence ID). In other words, high-confidence IDs will be highly accurate. By contrast, weaker memory signals that surpass only the leftmost criterion for making an ID with low confidence are almost as likely to be incorrect as correct (13% of target-present trials result in a correct low-confidence ID; 9% of target-absent trials result in an incorrect low-confidence ID). Thus, low-confidence IDs will be inaccurate according to this account. Although this is just one specific example, it illustrates why it has long been understood that a strong confidence-accuracy relationship is an inherent feature of signal-detection theory.

How Non-Pristine Testing Conditions Harm the Confidence-Accuracy Relation

Although signal-detection theory's prediction of a good confidence-accuracy relation is well founded, it tends to be based on an assumption that the only source of information for confidence is the strength of the memory signal. And, in a typical memory experiment, signal strength is the only available informational cue on which to base one's confidence. But eyewitness confidence in an

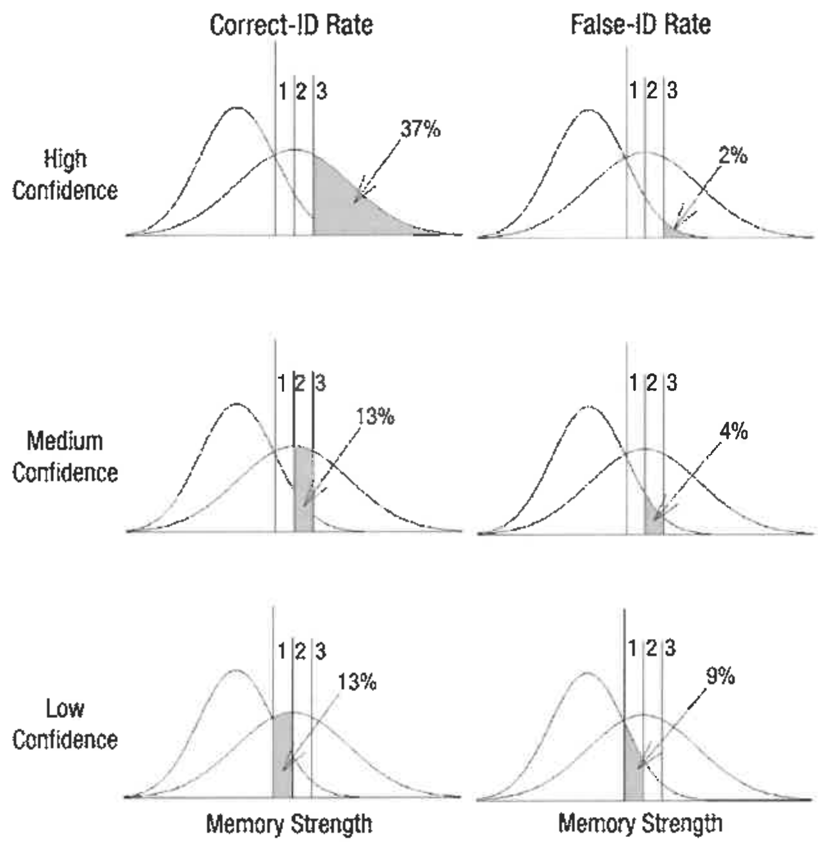


Fig. 11. Signal-detection-based interpretation of correct-ID rates (left panels) and false-ID rates (right panels) for high-confidence (top), medium-confidence (middle), and low-confidence (bottom) IDs. In each panel, the lure (innocent suspect) distribution is the narrow distribution on the left and the target (guilty suspect) distribution is the wider distribution on the right. Confidence criteria are shown as vertical lines, with the tallest vertical line representing the criterion for making an ID. The numbers 1, 2, and 3 represent low, medium, and high confidence, respectively.

identification is, in effect, the eyewitness's belief about the chances that the person he or she has identified is the perpetrator. And, as with other beliefs, eyewitnesses probably use whatever informational cues they have available to them when they state their confidence (Smalarz & Wells, 2015). If the only informational cue the eyewitness has at the time of making a confidence statement is a sense of the strength of the signal, then we would expect a good relation between accuracy and confidence because signal strength should be closely related to whether the target is the perpetrator or not. If, on the other hand, the eyewitness makes an identification and then overhears some seemingly confirmatory comment before making a confidence statement (e.g., "Your co-witness identified the same person"), then this confirmatory information is likely to be an additional cue driving his or her belief about the chances that the person identified is the perpetrator. In that case, the confidence of the witness is not based purely on the strength of the memory signal. If the confidence statement is based on considerations other than signal strength, then signal-detection theory's prediction of a confidence-accuracy relation no longer holds.

In our account, the requirement of pristine testing conditions applies not only to the composition of the lineup but also to the confidence statement, which should be assessed by the lineup administrator at the time of the initial identification (ideally, by a double-blind administrator whose behavior would not be influenced by knowledge of who the suspect is) before any other events can contaminate the confidence judgment. Consider, for example, the problem of assessing the confidence of an eyewitness who has been asked repeatedly to identify the same person (e.g., at the lineup, at a pretrial hearing, at trial). In such cases, the signal strength is likely to feel stronger to the eyewitness each time he or she encounters the person. Of course, the increase in signal strength is the result of repeated presentations of the suspect rather than the strength of the initial memory.

If, however, the witness fails to appreciate the effect of the intervening exposures on memory strength and relies on the (usually diagnostic) strong memory signal during subsequent tests, an error-prone high-confidence ID will be made. In the signal-detection model illustrated in Figure 11, this situation would be conceptualized as both distributions shifting to the right (as memory strength increases with repeated presentations) with the confidence criteria remaining fixed in place. In that case, high-confidence accuracy would plummet, because a much larger percentage of the lure distribution would now exceed the rightmost high-confidence criterion. Although a higher percentage of the target distribution would also now exceed the rightmost high-confidence criterion, the proportionate increase in false IDs would exceed the proportionate increase in correct IDs, so high-confidence accuracy would decrease. In effect, a source-monitoring failure will result in the witness relying on an internal memory cue—namely, strong memory—that is ordinarily diagnostic but no longer is (D. S. Lindsay, 2014; Roediger & DeSoto, 2015).

Why does a lineup that is composed of weak fillers (an unfair lineup) undermine our ability to infer high accuracy from high confidence? There are likely several reasons. First, one should not overlook the simple fact that unfair lineups increase the rate of mistaken identifications of innocent suspects at all levels of confidence. In a perfectly fair six-person lineup, for example, the maximum possible rate of mistaken identifications of an innocent suspect is 16.7%. And that maximum rate assumes that witnesses are performing at chance, that the perpetrator is never present in the lineup, and that all witnesses make an identification. But if just half of those witnesses do not make an identification and the perpetrator is in the lineup half of the time, the rate of mistaken identifications of innocent suspects from fair lineups would be less than 5% for low-confidence eyewitnesses and near floor for high-confidence eyewitnesses. An unfair lineup, in contrast, runs a much higher overall rate of mistaken identifications of the innocent suspect. This higher rate of mistaken suspect identifications from unfair lineups means that some are likely to end up in the high-confidence category.

In addition to raising the overall level of mistaken identifications of innocent suspects, there is also some evidence that unfair lineups can increase the confidence with which eyewitnesses make a mistaken identification. For example, as noted earlier, Charman et al. (2011) found that including highly dissimilar "dud" lineup members inflated witnesses' confidence in their mistaken identification of a non-dud. In a more recent study, Horry and Brewer (2016) manipulated the similarity between the suspect and the fillers in four-person simultaneous lineups and found that confidence judgments for positive identifications were predicted by the balance of evidence between the chosen item and the unchosen alternatives. In other words, as target-filler similarity decreased, confidence increased. This suggests that simultaneous lineup decisions are based at least in part on a relative memory-strength signal, which may be the reason why unfair lineups are so problematic. In an unfair lineup, the suspect (innocent or guilty) will generate a strong memory-match signal relative to those generated by the other lineup members (in Fig. 11, this would be conceptualized as both distributions being shifted to the right with the confidence criteria remaining fixed). The result would be a bias to choose that individual (Wells, 1984), even when making a high-confidence ID. As a bias to choose with high confidence increases, accuracy decreases. All of these problems are avoided (or at least minimized) if fair lineups are used.

An alternative but related theoretical interpretation is provided by fuzzy-trace theory's distinction between verbatim and gist memory traces (Brainerd & Reyna, 2005). According to fuzzy-trace theory, witnesses store both verbatim traces of the perpetrator plus more general (gist) traces of conceptually related information. Applied to eyewitness identification, the verbatim trace would be the perceptual representation of the perpetrator's face, whereas the gist trace might correspond to the general description of the perpetrator (e.g., an approximately 20-year-old White male with short dark hair and a scruffy beard). Depending on how retrieval is tested, a witness will rely on either the verbatim trace or the gist trace. When a participant relies on a verbatim trace, a strong memory-match signal (and attendant high confidence) will occur only when a face in the lineup matches that trace. As a general rule, such a match will occur only when the actual perpetrator is in the lineup. Thus, high-confidence accuracy will be high. The use of a pristine lineup seems well suited to promoting the retrieval of a verbatim trace because everyone in a fair lineup matches the gist (so the gist trace is of no help). However, in an unfair lineup, only the suspect corresponds to the gist trace, thereby promoting reliance on the gist trace instead of the verbatim trace. As noted by Brainerd and V. F. Reyna (2002):

> Retrieval of gist traces usually supports a more generic form of remembering, sometimes called familiarity, in which nonexperienced items are perceived to resemble experienced items but their occurrence is not explicitly recalled. However, when gist traces are especially strong, they can support high levels of phantom recollective experience for certain types of nonexperienced items—namely, items that are good cues for the gist of experience. (p. 166)

In other words, an unfair lineup might at times promote strong phantom recollection, leading to high-confidence errors. Accordingly, fuzzy-trace theory provides an additional theoretical rationale for recommending that the police use fair lineups.

General Conclusions

Our review of research concerned with the confidence-accuracy relationship in eyewitness identification is the first since Sporer et al. (1995) reviewed the literature more than 20 years ago. They found that when the analysis was limited to choosers, the correlation between confidence and accuracy was considerably higher than it was previously thought to be. That was their main message, even though their article is cited surprisingly often as suggesting the opposite. Nevertheless, the measure they used to assess that relationship—the point-biserial correlation coefficient—does not directly address the question of most interest to judges and juries. The point-biserial correlation coefficient is a perfectly reasonable effect-size statistic for a comparison between the average level of confidence associated with correct IDs versus the average level of confidence associated with incorrect IDs. However, the question asked by judges and juries concerns the average accuracy associated with suspect IDs made with a particular level of confidence. The correlation coefficient does not directly provide that information, but a calibration plot comes closer to doing so (Juslin et al., 1996). A calibration plot displays the proportion of correct IDs for choosers (or non-choosers) as a function of the level of confidence expressed, with confidence measured using a 100-point scale.

Calibration studies have consistently shown that for choosers, the confidence-accuracy relationship is strong (e.g., Brewer & Wells, 2006). The relationship is strong in the straightforward sense that high-confidence accuracy is much higher than low-confidence accuracy. Still, most calibration studies have found that highly confident witnesses are overconfident, and in one sense they are. Although CAC analysis treats only innocent-suspect IDs as relevant errors, from the eyewitness's point of view, filler IDs and innocent-suspect IDs are both relevant errors. Thus, when witnesses are asked to provide a confidence rating (e.g., 90%) that is commensurate with their accuracy (e.g., 90% of their IDs are guilty-suspect IDs, whereas 10% of their IDs are filler IDs or innocent-suspect IDs), their actual accuracy (e.g., 80% correct) can be said to reflect overconfidence. However, judges and juries in a case involving eyewitness-identification evidence are not interested in using an eyewitness's confidence to help them decide whether the witness picked a filler. Judges and juries already know that this particular witness did not pick a filler—the witness picked the suspect. Hence, judges and juries want to know how likely it is that the suspect is the perpetrator given that the witness identified the suspect with a particular level of confidence. The answer to their question, therefore, is provided by an analysis of the accuracy of suspect IDs per se without consideration of filler IDs. Of the suspect IDs that are made with a particular level of confidence, what proportion of those IDs were of guilty suspects and what proportion were instead of innocent suspects?

The answer to that key question is provided by CAC analysis, which is a measure of suspect-ID accuracy at each level of confidence for the base rate of target-present lineups used in the study (usually 50%). A more complete picture is provided by a Bayesian analysis that indicates what suspect-ID accuracy would be for the full range of possible base rates (0%–100%). Analyses of suspect-ID accuracy show that for a wide range of base rates, high confidence implies high accuracy (with no sign that witnesses are overconfident) and low confidence implies much lower accuracy. This is true of both lab studies and police department field studies, so long as pristine testing conditions are used. However, when the base rate is low enough (e.g., less than 25% of the lineups contain a guilty suspect), accuracy starts to become compromised across the board (even for high-confidence IDs). That fact provides a rationale for treating the base rate of guilty suspects as a system variable and for taking steps to ensure that the base rate is not unreasonably low. One way to do so is to require some objective evidence of guilt before placing a suspect in a lineup (Wells et al., 2015).

Importantly, a low-confidence ID on an initial test of memory from a lineup signals low accuracy whether or not pristine testing procedures are used. For this reason, low confidence should never be ignored and should instead always raise red flags about the reliability of the ID (Wixted et al., 2015). Although low-confidence IDs have some probative value when pristine procedures are used, under non-pristine testing conditions, they are even more error prone. As noted earlier, the majority of DNA exoneration cases in which eyewitness misidentification played a significant role were associated with, at best, a low-confidence ID on the initial memory test (Garrett, 2011). In some cases, the witness initially made a non-ID (i.e., confidence was so low that the witness identified no one) or a filler was identified. Thus, a low-confidence initial ID of a suspect from a lineup (or worse) corresponds to an uncomfortably high probability that the suspect is innocent. Had this simple fact been better understood by the legal system, many of the innocent defendants who were convicted based in part on a high-confidence ID that occurred in court may never have been convicted in the first place.

The news about the unreliability of a low-confidence initial ID will come as no surprise to most readers. Presumably, most readers are already under the impression that eyewitness memory is inherently unreliable, such that a suspect ID is error prone even under the best of conditions and even when confidence is high. Thus, the main news we have to offer is that eyewitness memory is not inherently unreliable. Under pristine testing conditions, a high-confidence suspect ID appears to be highly probative of guilt. Ignoring that fact—as the legal system is increasingly inclined to do—only serves to inappropriately exonerate the guilty. At the same time, ignoring low confidence at the time of an initial ID inappropriately imperils the innocent. The take-home message is that initial eyewitness confidence obtained from a pristine eyewitness-identification procedure serves both of the fundamental goals of the criminal justice system: to clear the innocent and to convict the guilty. By contrast, any later expression of confidence (including the confidence expressed by the eyewitness at trial in front of a jury) should be ignored, because doing otherwise works against the cause of justice.

Filler IDs and non-IDs are probative of innocence

Just as suspect-ID accuracy provides the information of interest to judges and juries tasked with evaluating the reliability of an eyewitness who has identified a suspect, analyses performed separately on filler IDs provide the information of interest to judges and juries tasked with evaluating the implications of the fact that an eyewitness picked a filler from a lineup instead of a suspect. Such an eyewitness would not testify against the defendant (because the eyewitness did not identify the defendant), but the fact that a filler ID occurred at an earlier stage of investigation nevertheless provides relevant information. The fact that a filler ID was made is somewhat probative of innocence. In other words, when filler IDs are examined separately, the data suggest that, given that a filler ID occurred, it is somewhat more likely that the lineup contained an innocent suspect than a guilty suspect.

In other cases, the eyewitness may have made a non-ID at the outset of the investigation. In a case like that, judges and jurors would be interested in the information value of a non-ID, and that information is provided by separately performed analyses of lab data for eyewitnesses who made non-IDs from target-present and target-absent lineups. When such an analysis is performed, the data indicate that non-IDs are also probative of innocence. The key point is that whether a case involves a suspect ID (the kind of ID that has helped to send innocent people to prison), a filler ID, or a non-ID, the information value of the ID in question is provided by analyzing the data separately, not by combining the data across suspect IDs, filler IDs, and non-IDs or by combining the data for choosers (suspect IDs and filler IDs) and analyzing them separately from data for non-choosers (non-IDs).

One of the relevant situations in which good records of rejections and filler IDs is important is in multiple-witness cases. Suppose, for example, that one witness identified the suspect and the other two rejected the lineup. What does that mean? Clark and Wells (2008) analyzed a large number of lab studies to estimate the probability that the suspect was the perpetrator under various combinations of suspect-ID, filler-ID, and lineup-rejection decisions in multiple-witness cases. In most cases, if one witness identified the suspect and the other two either rejected the lineup or picked a filler, the overall evidence pointed toward innocence rather than guilt of the suspect. Going forward, it will be important to address questions like this, taking into account IDs made with various levels of confidence (e.g., one high-confidence suspect ID and two low-confidence filler IDs).

Clark and Wells's (2008) analysis of the multiple-witness situation made it clear that one cannot ignore the witnesses who failed to pick the suspect. Nevertheless, in a 2012 national survey of U.S. law enforcement agencies, 37% of the agencies reported that they do not even write a report making a record of a lineup if the witness did not identify the suspect in the case (Police Executive Research Forum, 2013). Following on the lineups-as-experiments analogy described earlier in this article, this is akin to an experimenter ignoring data that are inconsistent with the hypothesis. Wells et al. (2015) argued that a failure to make a clear record of non-IDs and filler IDs could be construed as a "Brady violation"—that is, the violation of a constitutional requirement that the state reveal to the defense any evidence that might favor the defense (*Brady v. Maryland*, 1983).

Non-pristine testing conditions

How informative is confidence in a suspect ID that was made under non-pristine testing conditions? This is an important question to consider because, in the real world, pristine testing conditions will not always be achieved. Scientific research has clearly established that certain non-pristine testing conditions severely compromise the information value of eyewitness confidence. We consider them here.

Initial versus later confidence. Expressions of confidence by the eyewitness beyond the confidence statement at the initial identification are potentially problematic because a variety of factors (e.g., post-ID feedback) can inflate confidence without increasing accuracy. Thus, only an initial confidence statement—one that is made before there is much opportunity for confidence contamination to occur—provides reliable information. That fact underscores the importance of a recommendation long made by eyewitness-identification researchers and recently reiterated by the National Academy of Sciences committee: The initial confidence statement made by an eyewitness should be recorded and preserved. In this regard, another recommendation by the National Academy of Sciences committee—to videotape the witness-identification process—takes on special importance. Juries typically see an eyewitness make a high-confidence ID only in the courtroom, and they are heavily influenced by it. This is unfortunate because only the first ID, which occurred back at the beginning of the police investigation, provides diagnostic information about the reliability of the ID. With regard to its influence on jury decision making, an abstract discussion of the fact that confidence was low during an initial ID may have a hard time competing with the live expression of high confidence that occurs in the courtroom. However, if the initial lineup procedure were video recorded, jurors would have direct evidence that the eyewitnesses' initial level of confidence was low—evidence that would likely help them to understand that the ID is unreliable no matter what the witness now says.

Until relatively recently, video recording of all identification procedures was not practical for some jurisdictions because of the financial costs and video storage difficulties involved. Today, however, that is no longer true. Nevertheless, there are likely to be some cases in which witness cooperation is an issue. For example, if a witness who is critical to a case fears being video recorded (e.g., out of concern that the recording will end up on the Internet, where gang members will see who identified their comrade), then proceeding with the identification procedure without video recording it (perhaps instead audio taping it) might be advisable. Still, where possible (presumably in the large majority of cases), video recording the session will go a long way toward ensuring the integrity of the identification procedure and providing the jury with the information it needs about eyewitness confidence.

Having reliable information about the confidence of the eyewitness at the initial identification allows the defense to learn about and explain to the jury that confidence inflation has occurred. Some lab-based evidence has shown that, as one would hope, upon learning that a witness who was highly confident at trial was actually not confident at the time of the initial identification, mock jurors discounted their ratings of witness accuracy and the defendant's probability of guilt (Bradfield & McQuiston, 2004). On the other hand, Jones, Williams, and Brewer (2008) found that an "explanation" from the witness (e.g., "I was nervous at the time but now I am confident") led mock jurors to discount the low initial confidence of the witness and be more influenced by his or her later confidence. Indeed, we have concerns about how these problems would play out in pre-court and court proceedings to the extent that witnesses who were initially not confident would find some reason to explain away their initial lack of confidence and lead the court to rely on the inflated confidence that they had developed. One solution might be to adopt a hard-and-fast judicial rule stating that only the initial confidence of an eyewitness, made in good faith, is permissible in court. Another solution might be to adopt jury instructions stating that only confidence in an initial, good-faith attempt at an identification provides valid information about its reliability.

Fair versus unfair lineups. Another non-pristine testing condition that clearly compromises the information value of eyewitness confidence is an unfair lineup. Study after study has shown that if the innocent suspect in the lineup resembles the perpetrator to a greater extent than the fillers do (e.g., if the innocent suspect matches the description of the perpetrator more than the fillers do), high-confidence suspect-ID accuracy is greatly reduced (as illustrated earlier in Fig. 6). The importance of this observation is hard to overstate. If an unfair lineup is used, then the take-home message in this article does not apply. Mistakenly assuming that a high-confidence initial ID is highly accurate even when an unfair lineup is used is a recipe for wrongfully convicting the innocent.

Blind versus non-blind lineups. The blind lineup-administration procedure logically eliminates a potential source of error because the lineup administrator cannot possibly—intentionally or otherwise—steer the witness to the suspect in the lineup or provide post-ID praise to the witness for "getting it right" (thereby inflating even the initial statement of confidence). After an identification, even statements from a lineup administrator such as "you have been a really great witness" inflate the confidence of witnesses who have made a mistaken identification, but such statements do not inflate confidence if the witness knows that the lineup administrator is blind as to which lineup member is the suspect and which are fillers (Dysart, Lawson, & Rainey, 2012). In addition, there is evidence that lineup administrators influence witness confidence even when the administrators are given an unbiased script that they are supposed to follow (Garrioch & Brimacombe, 2001). Furthermore, lab data have shown that when people are assigned to the role of a lineup administrator, they tend to not make records of filler IDs when they know which lineup member is the suspect (non-blind lineup administrators), but they faithfully make such records when they do not know the status of the identified lineup member (blind lineup administrators; see Rodriguez & Berry, 2014). These considerations explain why blind lineup administration has long been recommended by eyewitness-identification researchers and why that recommendation was also recently endorsed by the National Academy of Sciences committee.

The point is that (a) confidence is a reliable indicator of accuracy under pristine testing conditions; (b) confidence is a much less reliable indicator of accuracy under certain non-pristine testing conditions (e.g., when an unfair lineup is used or when the test is not the initial ID test); and (c) eyewitness expressions of confidence can be influenced by non-blind lineup administrators, which is an undesirable outcome no matter what its effect on accuracy might be. Obviously, confidence may or may not be a reliable indicator of accuracy under other conditions that have not yet been subjected to scientific investigation. Later, we recommend some research priorities for further investigating the eyewitness confidence-accuracy relationship.

Estimator variables and confidence in a suspect ID

In the studies reviewed here, eyewitnesses who were tested using pristine procedures appropriately adjusted their confidence downward when they were aware that no one in the lineup strongly matched their memory of the perpetrator. This is just another way of saying that there is a strong relationship between confidence and accuracy. That finding may have some non-obvious but nevertheless important implications for how people generally think about the effect of various estimator variables on eyewitness-identification accuracy. Consider, for example, how juror guidelines in Massachusetts instruct juries to think about estimator variables. Those instructions list a variety of factors than can make memory worse, on average (e.g., long retention interval, short exposure time, stress, the presence of a weapon), and they invite jurors to believe that if one or more of those factors is present, then the reliability of the ID should be regarded as less trustworthy than it otherwise would be. As intuitively appealing as this line of thinking might be, the evidence suggests that it may not be valid.

To illustrate this point, we consider the fact that a long retention interval typically results in worse overall memory performance compared to a short retention interval. Does that fact imply that a high-confidence initial ID of a suspect made after a long retention interval is less trustworthy than a high-confidence initial ID of a suspect made after a short retention interval? Not necessarily. Eyewitnesses have a sense of how well each lineup member matches their memory, and if the memory is weak, they are not likely to have high confidence. That is, as memory fades with the passage of time, eyewitnesses will be less likely to experience a strong memory-match signal when viewing the members of a photo lineup. As a result, witnesses might make more errors but, critically, those errors are likely to be associated with low confidence (because high-confidence IDs are typically made when the memory-match signal is strong, not when it is weak, as it generally would be following a long retention interval). Nevertheless, for the smaller percentage of eyewitnesses who do make a high-confidence ID despite a long retention interval, their average accuracy could be every bit as high as that for the larger percentage of eyewitnesses who make a high-confidence ID following a short retention interval.

Although additional research is certainly needed, the available evidence indicates that eyewitnesses may often appropriately adjust confidence to the prevailing memory conditions, contrary to Deffenbacher's optimality hypothesis (Deffenbacher, 1980). Palmer, Brewer, Weber, and Nagesh (2013, Experiment 1) compared immediate versus 1-week-delayed performance in a large-scale experimentally controlled field study. Not surprisingly, they reported that overall accuracy was lower following the 1-week retention interval than on the immediate test, but as shown in Figure 4l, the accuracy of high-confidence IDs was equally high either way. The same was true when overall memory strength was manipulated by varying exposure duration from 5 seconds to 90 seconds (also shown in Fig. 4l) or by varying whether or not attention was distracted during exposure (Fig. 4m). In each case, overall memory performance was weaker in one condition compared to the other, but high-confidence accuracy was the same either way. With regard to a retention-interval manipulation, Juslin et al. (1996), Read et al. (1998), and Sauer, Brewer, Zweck, and Weber (2010) all reported a similar outcome (Figs. 4h, 4n, and 4p, respectively). Note that the Read et al. (1998) results are noteworthy in that those authors used retention intervals as long as 9 months.

Similar effects are evident for several other estimator variables. For example, Carlson and Carlson (2014) and Carlson, Dias, Weatherford, and Carlson (in press) found that although the presence of a weapon clearly led to worse memory performance overall (the weapon-focus effect), it had virtually no effect on the accuracy of identifications made with high confidence (Figs. 4c and 4d). The same outcome was observed by Dodson and Dobolyi (2016) for same-race versus cross-race IDs (Fig. 4f). Cross-race IDs were associated with significantly lower recognition memory performance compared to same-race IDs, but high-confidence IDs were highly (and similarly) accurate either way.

If these results generalize to the real world, they suggest that these estimator variables may not be particularly relevant to the reliability of an initial ID made with high confidence. Although definitive conclusions cannot yet be drawn, the overall pattern of results suggests that under pristine testing conditions, estimator variables that have long been thought to compromise the reliability of a suspect ID may not do so (because eyewitnesses appropriately adjust their confidence under poorer estimator-variable conditions). Still, it would be premature to make a definitive statement regarding the effect of different estimator variables on the accuracy of IDs made with high confidence because the issue has only recently been addressed using CAC analysis. In addition, a study by Lampinen, Erickson, Moore, and Hittson (2014) investigated the effect of distance on identification accuracy. This study used an old/new recognition procedure (not a lineup) in which each witness made 16 recognition decisions. Thus, its design was far removed from the kind of forensically relevant lineup designs that we have considered here. Nevertheless, it is worth noting that according to our estimates based on the ROC data presented in their Figure 4, high-confidence accuracy was always below 90% and became noticeably worse as distance increased, falling to approximately 70% correct at the longest distances tested. Whether the same would be true for lineups is unknown, but this result underscores the fact that more work is needed to determine the effect of estimator variables on high-confidence accuracy.

Mistaken-ID rates at the level of the lineup versus the courtroom: The plea effect

At this point it is important to note that we cannot necessarily assume that the chances that a high-confidence ID is mistaken at the level of the lineup are the same as the chances that a high-confidence ID is mistaken at the level of a trial. One reason, although not the only reason, is that guilty pleas (which do not go to trial) will remove many more guilty than innocent people from trials. This *plea effect*, originally described by Wells, Memon, and Penrod (2006), yields a distribution of innocent and guilty individuals at trial that is different from the distribution at the level of the lineup.

Let us assume that witnesses who were tested using pristine procedures (fair lineup, double-blind administrator, confidence measured at time of ID, etc.) and were 95% to 100% confident have a 98% chance of being accurate. In other words, only 2% of these witnesses would be mistaken. Suppose now that we have a defendant on trial who was identified by an eyewitness who made the identification under pristine testing conditions and was 95% to 100% confident. Can we assume, in the absence of any other evidence, that at the trial level there is only about a 2% chance that the person the witness identified is an innocent person? The answer is "not necessarily," especially in the U.S. legal system. Depending on its size, the plea effect could create a situation in which the chance that the defendant is innocent is much higher than 2%.

The plea effect (Charman & Wells, 2007; Wells et al., 2006) refers to the fact that most criminal convictions never involve a trial at all but instead are obtained through guilty pleas. In fact, over 95% of criminal convictions in the United States are attained through plea deals and are never brought to trial (Ross, 2006). Because fewer than 5% of felony convictions come from people who claim innocence and choose to take their case to trial, those who do so represent a small subset of defendants. And, although innocent people sometimes plead guilty (e.g., over 20% of the DNA exoneration cases involved an innocent person who pled guilty), it seems reasonable to assume that the chances that an innocent person would take a case to trial rather than plead guilty is much greater than the chances that a guilty person would take a case to trial.

Consider 10,000 suspect IDs made with high confidence. For the sake of simplicity, let's assume that all 2% of those who were mistakenly identified with high confidence ($10{,}000 \times 0.02 = 200$ innocent suspects) are prosecuted and take their case to trial (after all, they are innocent). And, let's assume that of the 98% who were accurately identified with high confidence ($10{,}000 \times 0.98 = 9{,}800$ guilty suspects), 97% ($9{,}800 \times 0.97 = 9{,}506$) take a plea and 3% ($9{,}800 \times 0.03 = 294$) instead go to trial. If this were the case, and if jury trials always resulted in guilty verdicts, then $100\% \times 9{,}506 / (9{,}506 + 200) = 95.1\%$ of guilty verdicts would arise through plea deals. Moreover, among those who took their case to trial (200 innocent suspects and 294 guilty suspects), the chances of the defendant being guilty based on the eyewitness-identification evidence alone would be slightly less than $100\% \times 294 / (294 + 200) \approx 60\%$. In other words, what is a mere 2% mistaken-identification rate at the level of the lineup becomes a 40% chance of innocence among cases that make it to trial. That reduction in accuracy at trial is, of course, offset by an increased level of accuracy associated with high-confidence IDs that ended in a plea deal instead of going to trial. In this example, because all of the innocent suspects went to trial, 100% of the defendants who were identified with high confidence and who accepted a plea bargain would be guilty.

Obviously, these numbers will change depending on the assumptions that are made. For example, instead of being equally likely to be forwarded for prosecution (as assumed in the example above), guilty suspects may be more likely than innocent suspects to be forwarded for prosecution. This might occur because guilty suspects are more likely to have independent corroborating evidence against them compared to innocent suspects. In addition, the 95% plea rate, which is based on all cases, may be an overestimate for eyewitness-identification cases because defense attorneys might believe that they have a better chance of acquittal in cases involving eyewitness-identification evidence than in many other types of cases. If we assume that suspects who have been identified with high confidence are twice as likely to be forwarded for prosecution if they are guilty than if they are innocent (because of a disparity in corroborating evidence), that 25% of guilty suspects choose jury trials (in hopes of discrediting eyewitness evidence), and that 50% of jury trials end in guilty verdicts, then 85% of all guilty verdicts in cases involving eyewitness identification would arise from plea bargains, and high-confidence ID accuracy at trial would be 96% correct.

Although the precise numbers cannot be known, it is important to appreciate that the plea effect changes the ratio of the innocent to the guilty among those who actually go to trial. The more the plea effect increases the ratio of the innocent to the guilty at trial, the less trustworthy a high-confidence ID becomes at trial (and the more trustworthy a high-confidence ID becomes for those who choose to accept a plea bargain).

The distinction between eyewitness-identification accuracy at the level of the lineup and eyewitness-identification accuracy at the level of cases that go to trial is important. An eyewitness expert giving trial testimony, for example, should be careful to not equate the mistaken-identification rate at the level of the lineup with the chances that the defendant is guilty in a particular case that made it to trial. A similar caution applies to the base-rate issue discussed previously (i.e., a high-confidence accuracy score estimated from a study that used a 50% target-present base rate does not directly apply to a jurisdiction that might have a much lower base rate). At the same time, these considerations do not undermine the general conclusion of the current article, namely that high-confidence eyewitness identifications made using pristine testing procedures have a very low rate of error.

Priorities for future research

How to collect a confidence statement from an eyewitness. Although confidence in an initial ID is highly predictive of accuracy, no police department field study has specifically investigated different methods for recording initial confidence. Should a confidence statement be taken in the witness's own words (as in Klobuchar et al., 2006), or should confidence be recorded using an explicit 3-point rating scale (as in Wixted et al., 2016)—or should a 100-point scale be used? Given the clear information value of initial confidence, this issue seems important to pursue.

How to create a fair lineup. Unfair lineups seriously degrade the information value of eyewitness confidence. One way to minimize the chances of creating an unfair lineup is to ensure that every member of the lineup matches the description of the perpetrator provided by the witness. However, this is a subjective process, and even an investigator who is trying to follow that directive might unintentionally create an unfair lineup. Indeed, in one condition of a recent police department field study (the blinded condition in Wixted et al., 2016), the lineups assessed by mock witnesses were found to be unfair in that the suspect in the lineup was selected, on average, more than the fillers based solely on the description. But even when care is exercised to make sure that all fillers match the description of the perpetrator that was provided by the witness, the lineup might not be fair. This is because eyewitnesses' verbal descriptions of perpetrators are often vague or incomplete, and sometimes the description does not even match the suspect (Luus & Wells, 1991). Some have proposed that the fillers should be matched to the suspect on major physical characteristics rather than just those contained in the eyewitness's description of the perpetrator (e.g., Lindsay, Martin, & Webber, 1994). Others have proposed that the fillers be selected based on their overall similarity to the suspect (Clark & Tunnicliff, 2001). Some have found that it is possible to make the fillers too similar to the suspect (which protects innocent suspects but reduces the chances of identifying perps; see Wells et al., 1993). And, as discussed earlier in this article, when someone becomes a suspect based on similarity to a composite or a surveillance image, simply matching fillers to the eyewitness's verbal description of the suspect is not sufficient. Clearly, the general idea that poor lineup fillers place innocent suspects at risk and confound our ability to rely on confidence is not in question, and we see evidence of this in the CACs shown in Figure 6. But there is a need to articulate more precisely what the criteria should be for making lineups fair. What tools can be developed for officers who are tasked with creating a lineup to make their job easier and more objective?

The effect of estimator variables on confidence. An important goal for future research will be to determine if the conclusions discussed above with respect to estimator variables apply to other estimator variables that are relevant to eyewitness IDs in the real world (e.g., high stress vs. low stress). The fact that estimator variables have an effect on overall memory accuracy is beyond dispute; what remains unknown is what effect they have on the confidence-accuracy relationship when the data are subjected to CAC analysis. This is an important issue to specifically investigate because variables that impair overall memory accuracy do not necessarily have any effect on the accuracy of suspect IDs made with high confidence (instead, they may affect only the frequency of high-confidence suspect IDs).

Exploring other ways of sorting between guilty and innocent suspects. The standard approach to assessing eyewitness-identification confidence is to ask the eyewitness how confident she or he is in the identification that was made. But research by Sauer, Brewer, and Weber (2008) found that collecting a witness confidence statement for each lineup member (rather than only the one who was chosen) provided a more informative indicator of recognition. Following on this finding, more recent research has shown promising results for procedures in which eyewitnesses do not pick someone out of a lineup at all but instead make a confidence judgment about whether each lineup member is the perpetrator (e.g., Brewer, Weber, Wootton, & Lindsay, 2012; Sauer, Brewer, & Weber, 2008; Sauer, Brewer, & Weber, 2012) or rate how well each face matches their memory of the perpetrator (Sauer, Weber, & Brewer, 2012). Results from profile analyses and classification algorithms have shown that such methods may be superior to the traditional eyewitness-identification task. Other work has examined decision time and shown that eyewitnesses make accurate identifications consistently faster than they make mistaken identifications (e.g., Dunning & Perretta, 2002; Sauer, Brewer, & Wells, 2008; Sporer, 1993). Our point here is simply that we do not want to close off the possibility that there might be other approaches to assessing the probability of a suspect's guilt that work even better than traditional methods.

Conclusion

According to the available data, the relationship between confidence and accuracy for an initial ID from an appropriately administered lineup is sufficiently impressive that it calls into question the very notion that eyewitness memory is generally unreliable. Eyewitness memory can certainly *become* unreliable as a result of influences introduced by the legal system (feedback, repeated exposure to the suspect, misinformation, biased lineup composition, etc.), but the same is true of any kind of evidence, including DNA evidence. A contaminated eyewitness memory test, like a contaminated DNA test, is not reliable. However, the available research suggests that when pristine testing procedures are used, an initial ID made with high confidence is highly indicative of accuracy. Perhaps even more importantly, an initial ID made with low confidence—whether testing conditions are pristine or not—is highly error prone. A better appreciation of that simple fact might have prevented most of the DNA exonerees from being convicted in the first place. Thus, instead of disregarding eyewitness confidence altogether, the legal system should draw a distinction between initial confidence that was obtained using pristine testing procedures and confidence obtained later or under conditions known to compromise the confidence-accuracy relationship.

Appendix A

An illustration of a strong confidence-accuracy relationship despite a low point-biserial correlation

Twenty years ago, Juslin, Olsson, and Winman (1996) explained that the point-biserial correlation coefficient is problematic for assessing the confidence-accuracy relationship because its value can be low even when eyewitnesses exhibit perfect calibration (such that 100% confidence implies 100% accuracy, 90% confidence implies 90% accuracy, etc.). However, they did not illustrate what the point-biserial correlation coefficient actually measures, nor did they reanalyze any of the prior data to show what those data look like when analyzed in a more appropriate way. This may account for why, to this day, scientists continue to rely on the point-biserial correlation coefficient to measure the relationship between confidence and accuracy and why the legal system does so as well. Here, we explain what this statistic actually measures and why it should no longer be used if the goal is to inform the legal system about the reliability of a suspect ID made with a particular level of confidence. Again, it is a perfectly valid statistic when used for other purposes (Rosnow, Rosenthal, & Rubin, 2000), and it does signal a strong relationship between confidence and accuracy when its value is high (D. S. Lindsay, Nilsen, & Read, 2000; D. S. Lindsay, Read, & Sharma, 1998). However, for the purpose of predicting eyewitness-identification accuracy from an eyewitness's expression of confidence, it can be misleading because it does not necessarily indicate a weak relationship between confidence and accuracy when its value is low (as has been assumed by researchers and the legal system alike).

Table A1 presents hypothetical data generated by 30 "choosers" who have made an ID from a lineup and rated confidence using a 5-point confidence scale (1 = *low confidence*, 5 = *high confidence*). Choosers make one of four possible decisions: correctly identifying a suspect from a target-present lineup, incorrectly identifying a suspect from a target-absent lineup, incorrectly identifying a filler from a target-present lineup, or incorrectly identifying a filler from a target-absent lineup. Thus, all of the

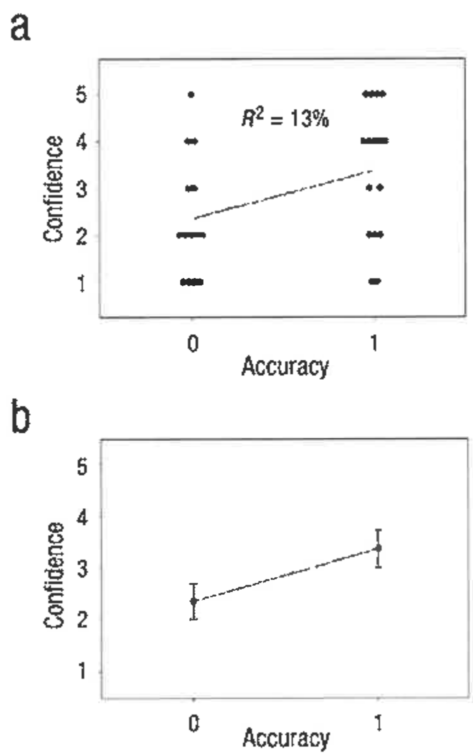


Fig. A1. Hypothetical data generated by 30 "choosers" who have made an ID from a lineup and rated confidence using a 5-point scale, with confidence plotted as a function of accuracy. In panel (a), each point is a data point from one participant, and the line is the line of best fit. In panel (b), the points represent the data averaged across confidence, and the bars represent standard errors.

accuracy scores of "1" in Table A1 correspond to suspect IDs from target-present lineups (which is the only correct response for a chooser). For simplicity, these hypothetical data are conceptualized as having come from a mock-crime study in which it is known whether the suspect in the lineup is innocent or guilty (e.g., as illustrated in Fig. 1). The data in Table A1 have been chosen to illustrate a point about the correlation between confidence and accuracy, not to reflect what typical data necessarily look like. The correlation between the 30 accuracy scores and the 30 corresponding confidence scores shown in the two rightmost columns of Table A1 is .36, which is slightly lower than the generally accepted value of .41 for choosers.

Figure A1a illustrates the fact that computing a point-biserial correlation coefficient is tantamount to fitting a straight line through the data when confidence is plotted as a function of accuracy coded in binary format (0 = inaccurate, 1 = accurate). Each point represents one participant, and the points for different participants that would fall atop one another have been slightly spread out on the accuracy dimension to show how many participants are associated with each confidence-accuracy pair. The best-fitting line is the one that minimizes the sum of the squared deviations (vertically) between the line and the 30 individual data points. It is difficult to imagine how judges and juries could extract useful information about the likely reliability of a particular suspect ID (e.g., one made with high confidence) from data analyzed in this manner.

Figure A1b shows the same data except that the confidence ratings have been averaged together to make a more interpretable graph. This figure clearly shows that the average level of confidence is higher for correct IDs than for incorrect IDs. In fact, this is how the data were plotted in Figure 1 of Sporer, Penrod, Read, and Cutler's (1995) seminal article. When the data are analyzed in this manner, the result is presumably more interpretable to judges and juries. However, a problem with Figure A1b is that it plots the unaveraged dependent measure (accuracy coded as 0 or 1) on the *x*-axis and the averaged predictor variable (confidence) on the *y*-axis. This would be the appropriate way to plot the data if you knew, for each eyewitness, whether his or her ID was correct or incorrect and wanted to estimate his or her likely level of confidence. If that were the question of interest, then the point-biserial correlation coefficient would be a reasonable effect-size statistic to help conceptualize the results of a *t* test (for example) comparing average confidence for correct decisions versus average confidence for incorrect decisions (Rosnow et al., 2000). Yet this is not the question of interest, because in actual practice, the situation is reversed: An eyewitness provides a confidence rating associated with an ID (this is the predictor variable, which is not averaged), and the legal system wants to make the best estimate as to the likely accuracy of that ID (this is the dependent variable, and it equals the average level of accuracy associated with each level of confidence that an eyewitness might express). This logic suggests, as Juslin, Olsson, and Winman (1996) pointed out, that plotting average accuracy (on the *y*-axis, as the dependent measure) versus different levels of confidence (on the *x*-axis, as the independent measure) is the sensible way of representing the data and addressing the question of interest. Only when plotted this way are the data presented in a manner that provides an answer to the critical question asked by the criminal justice system: Given that an eyewitness has a particular level of confidence in his or her ID, how accurate is that ID likely to be?

Figure A2a shows the same data plotted in Figure 2a except that the axes have been reversed to plot the independent variable (confidence) on the *x*-axis and the dependent variable (accuracy) on the *y*-axis. Obviously, because the information in Figures A1a and A2a is the same, the best-fitting line corresponds to the same point-biserial correlation coefficient (.36) as in Figure A1a. However, even with the variables appropriately reversed, the data do not yet provide much in the way of useful

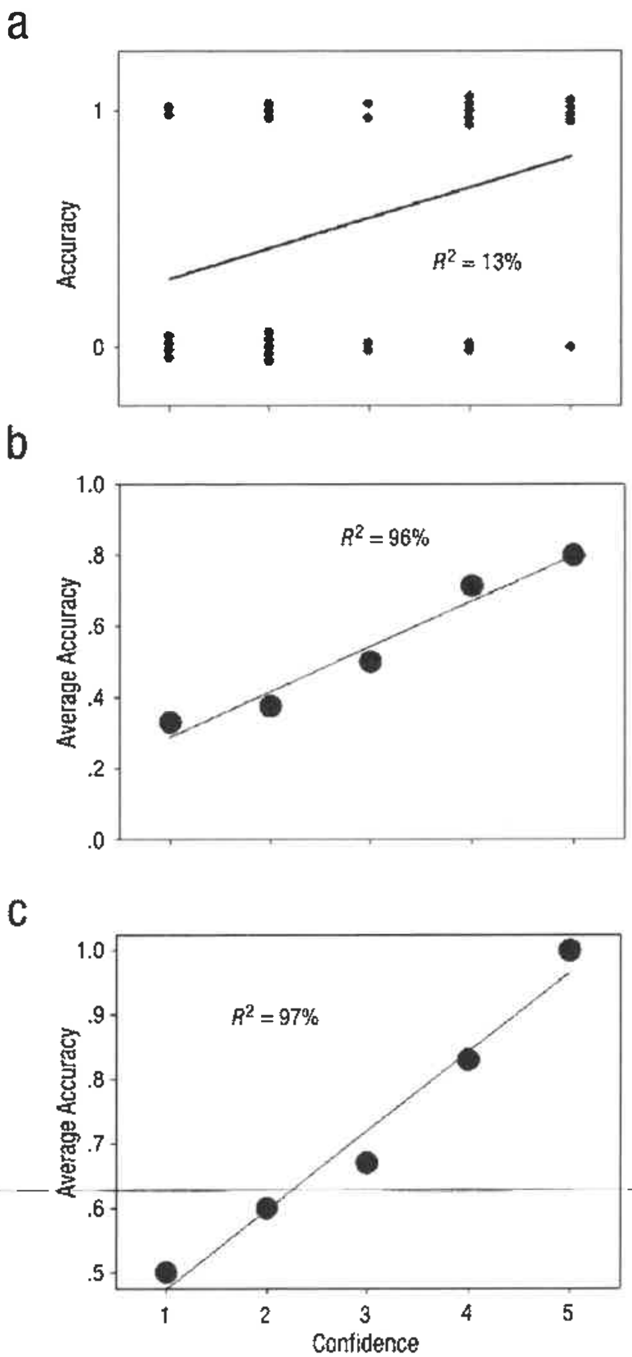


Fig. A2. Hypothetical data generated by 30 "choosers" who have made an ID from a lineup and rated confidence using a 5-point scale, with accuracy plotted as a function of confidence. In panel (a), each point is a data point from one participant, and the line is the line of best fit. In panel (b), the points represent the data averaged across accuracy. Panel (c) shows the confidence-accuracy characteristic curve for when only suspect IDs are included in the accuracy calculation.

information to courts of law. Figure A2b shows the same data as Figure A2a except that the binary accuracy scores associated with each level of confidence have been averaged together. Now the data are depicted in a way that is useful to judges and juries. How accurate is an ID made with the highest level of confidence (a rating of 5) likely to be? How accurate are medium-confidence IDs (e.g., ratings of 3)? And how accurate are low-confidence IDs (e.g., ratings of 1)? The answers to these questions are meaningful to judges and jurors (Mickes, 2015), and all of this information is available in Figure A2b. By contrast, the point-biserial correlation coefficient (obtained by fitting the data in Figures A1a and A2a with a straight line) does not provide this information.

For these hypothetical data, which yield a point-biserial correlation of .36, IDs made with high confidence (a rating of 5) are 80% correct, whereas IDs made with low confidence (a rating of 1) are only 33% correct. Thus, a point-biserial correlation that is even less than the magnitude of the widely accepted estimate for choosers (i.e., .41) is consistent with high-confidence IDs being far more accurate than low-confidence IDs. But even this improved analysis underestimates the reliability of eyewitness identification for the same reason that the calibration curves do. What is the problem?

Of the 30 hypothetical choosers shown in Table A1, 22 picked a suspect and the other eight picked a filler (as indicated in Column 2). Imagine that in none of these 22 cases is there any incriminating evidence against the suspect other than the evidence that might be provided by the eyewitness. In this example, eight eyewitnesses chose a filler, thereby ending any further consideration of the suspects in those lineups. But 22 of them identified a suspect, and those 22 identifications are the ones that would go forward as direct evidence of the suspect's guilt. Some of these identifications involve a suspect ID made with high confidence and others involve a suspect ID made with low confidence. The judges and juries in those cases would be interested in knowing whether or not such IDs are reliable. Stated differently, their question is as follows: Of the eyewitness-identification cases that end up before judges and juries (which are limited to identified suspects), what does confidence tell us about the reliability of the ID? Note that this is a question about the 22 cases that go forward to the prosecution using eyewitness identification as direct evidence of the suspect's guilt, not about the full set of 30 cases involving choosers. The answer to this question is provided by limiting the analysis not just to choosers but to *choosers who identify a suspect*—just as the legal system limits its consideration to choosers who identify a suspect.

Table A2 presents the hypothetical data from the 22 choosers who identified a suspect (i.e., it presents the data that would be of interest to judges and jurors), and it now highlights the six choosers in this example who incorrectly identified an innocent suspect from a target-absent lineup. Although none of those six choosers identified an innocent suspect with high confidence (i.e., with

Table A1. Hypothetical Confidence-Accuracy Data From 30 Participant-Witnesses

Witness	Pick type	Lineup type	Accuracy	Confidence
1	Suspect	TP	1	4
2	Suspect	TP	1	4
3	Suspect	TA	0	1
4	Suspect	TP	1	1
5	Suspect	TA	0	2
6	Suspect	TP	1	4
7	Filler	TA	0	5
8	Suspect	TP	1	3
9	Filler	TA	0	2
10	Filler	TP	0	2
11	Suspect	TP	1	5
12	Suspect	TP	1	4
13	Filler	TA	0	1
14	Suspect	TP	1	4
15	Filler	TP	0	1
16	Filler	TA	0	2
17	Suspect	TP	1	2
18	Suspect	TP	1	5
19	Filler	TA	0	4
20	Suspect	TP	1	2
21	Suspect	TP	1	3
22	Suspect	TA	0	3
23	Suspect	TP	1	5
24	Suspect	TP	1	1
25	Suspect	TA	0	1
26	Suspect	TA	0	4
27	Suspect	TP	1	5
28	Suspect	TP	1	2
29	Filler	TA	0	3
30	Suspect	TA	0	2

Table A2. Hypothetical Confidence-Accuracy Data From the 22 Choosers From Table A1

Witness	Pick type	Lineup type	Accuracy	Confidence
1	Suspect	TP	1	4
2	Suspect	TP	1	4
3	Suspect	TA	0	1
4	Suspect	TP	1	1
5	Suspect	TA	0	2
6	Suspect	TP	1	4
8	Suspect	TP	1	3
11	Suspect	TP	1	5
12	Suspect	TP	1	4
14	Suspect	TP	1	4
17	Suspect	TP	1	2
18	Suspect	TP	1	5
20	Suspect	TP	1	2
21	Suspect	TP	1	3
22	Suspect	TA	0	3
23	Suspect	TP	1	5
24	Suspect	TP	1	1
25	Suspect	TA	0	1
26	Suspect	TA	0	4
27	Suspect	TP	1	5
28	Suspect	TP	1	2
30	Suspect	TA	0	2

Note: Confidence ratings range from 1 (*low confidence*) to 5 (*high confidence*). Accuracy is coded as 0 for inaccurate and 1 for accurate. TP = target present; TA = target-absent.

a rating of 5), four of the other 16 witnesses did identify a guilty suspect with high confidence (Witnesses 11, 18, 23, and 27). Thus, high-confidence suspect-ID accuracy in this hypothetical example is perfect (4 correct, 0 incorrect).

Figure A2c shows the results of this analysis when the data are limited to the 22 choosers in Table A1 who identified a suspect. Obviously, the relationship between confidence and accuracy for these hypothetical data is still very strong, in the sense that high-confidence IDs are far more accurate than low-confidence IDs (as illustrated in Fig. 3b). High-confidence suspect IDs are 100% accurate, whereas low-confidence suspect IDs are only 50% accurate (close to chance). Thus, not only is confidence highly diagnostic of accuracy, high-confidence suspect IDs in this hypothetical example are extremely accurate (as accurate as they could possibly be). Keep in mind that these are the very same data that when analyzed using the point-biserial correlation coefficient and including choosers who identify fillers (as in Fig. A1a) yield a value of .36. Even when the point-biserial correlation coefficient is computed for choosers who made suspect IDs (i.e., even when computed using the data in Table A2), its value is only .39. Thus, the correlation coefficient does not convey the information of interest to judges and juries. The data shown in Figure A2c do.

Figure A2c shows a confidence-accuracy characteristic curve (Mickes, 2015). Such a curve plots suspect-ID accuracy as a function of confidence that has been assessed using any numerical scale (in this example, a 1-to-5 scale). Suspect-ID accuracy is computed separately for each level of confidence, *c*, and is computed from the number of suspect IDs from target-present (*TP*) lineups, $nSID_{TP\text{-}c}$, and the number of suspect IDs from target-absent (*TA*) lineups, $nSID_{TA\text{-}c}$. More specifically, suspect ID accuracy for a given level of confidence is equal to $nSID_{TP\text{-}c}$ / ($nSID_{TP\text{-}c}$ + $nSID_{TA\text{-}c}$). In the example above, for high-confidence IDs (i.e., $c = 5$), $nSID_{TP\text{-}5} = 4$ and $nSID_{TA\text{-}5} = 0$, so high-confidence suspect-ID accuracy is 4 / (4 + 0) = 1.0. This accuracy score differs from the usual dependent measure in calibration studies, in which filler IDs are included in the denominator (as in Fig. A2b). Obviously, including filler IDs lowers the estimated accuracy score, although in this case it has little effect on the overall correlation between confidence and accuracy. However, the correlation is not relevant for what judges and juries want to know, because the correlation could be perfect and yet high-confidence IDs could still (hypothetically) be only 60% accurate. Hence our concentration on the confidence-accuracy characteristic and the probability correct associated with high- and low-confidence suspect IDs.

Appendix B

Estimating suspect-ID accuracy from a calibration score

Most of the calibration studies we reviewed did not present their data in sufficient detail to directly calculate suspect-ID accuracy, so we computed an estimate from the calibration data reported in figures. WebPlotDigitizer (http://arohatgi.info/WebPlotDigitizer/) was first used to estimate C_c (proportion correct, C, for each level of confidence, c). We then converted those scores, which included filler IDs, to scores that included only suspect IDs. The conversion from C_c to suspect-ID accuracy, $p(TP \mid SID_c)$, is straightforward. Using the most common calibration formula (which excludes filler IDs from target-present lineups), calibration for a given level of confidence is:

$$C_c = \frac{nSID_{TP-c}}{nSID_{TP-c} + nFID_{TA-c}} \qquad (1)$$

To convert C_c to suspect-ID accuracy, we use the following formula:

$$p(TP \mid SID_c) = \frac{c_c}{c_c + (1 - c_c)/n} \qquad (2)$$

where n = lineup size. As an example, imagine a study using eight-person lineups in which there were 80 correct high-confidence suspect IDs from target-present lineups and 80 high-confidence incorrect IDs from fair target-absent lineups that did not have a designated innocent suspect. Thus, $nSID_{TP\text{-}high} = 80$ and $nFID_{TA\text{-}high} = 80$. In that case, calibration for high-confidence IDs (Equation 1) would equal 80 / (80 + 80) = .50. However, to compute suspect-ID accuracy, the number of high-confidence filler IDs from target-absent lineups, $nFID_{TA\text{-}high}$, is divided by lineup size to estimate the number of innocent-suspect IDs from target-absent lineups, $nSID_{TA\text{-}high}$, where $nSID_{TA\text{-}high} = nFID_{TA\text{-}high} / n$. Note that suspect-ID accuracy is given by:

$$\frac{nSID_{TP-c}}{nSID_{TP-c} + nSID_{TA-c}}$$

Thus, for this example, suspect-ID accuracy (the proportion of suspect IDs that were correct) is 80 / (80 + 80 / 8), which reduces to 1 / (1 + 1 / 8) = .89. However, all we have is the reported calibration accuracy score of .50 (estimated from a figure). Using the above formula (Equation 2), the calibration score is converted into a suspect-ID accuracy score by computing .50 / [.50 + (1 − .50) / 8)], which reduces to 1 / (1 + 1 / 8) = .89. Thus, Equation 2 gives us the right answer (i.e., the same answer we came up with by directly computing suspect-ID accuracy from the raw counts of suspect IDs and filler IDs—the kind of information we do not have access to in most studies). Equation 2 was used to compute suspect-ID accuracy from the calibration scores for each level of confidence—scores that were estimated from the reported figures. All of the studies involved a base rate of approximately 50% (i.e., 50% of the lineups were target-present lineups, and 50% were target-absent lineups).

Appendix C

A short primer on base rates in lineups

The probability that some proposition is true (e.g., that a suspect is guilty) given the result of an evidentiary test (e.g., identification by a witness in a lineup test) is a function of both the diagnostic value of the evidence (e.g., the reliability of the identification) and the base-rate (or prior) probability that the proposition is true. This is often counterintuitive, and people commonly assume that the probability that a proposition is true is equal to the diagnostic value of the evidence without regard to the base rate. Consider, for example, a prostate exam that gives a positive result 98% of the time when there is cancer (a 98% hit rate) and a positive result only 2% of the time when there is no cancer (a 2% false-positive rate). Armed with such information, most people will assume that a positive result indicates a 98% chance of cancer. But that would be true only if one were sampling from a population of men for whom the base rate of prostate cancer was 50% to begin with. Suppose, however, the test is conducted on relatively young men for whom the base rate for prostate cancer is a mere 1%. In the 1%-base-rate population, a positive test result would yield a probability of cancer of slightly less than 5%, not 98%.

The influence of base rates is somewhat counterintuitive, but the math is not particularly difficult. Consider, for example, that in the 1%-base-rate population of young males, 999 of every 1,000 males tested would not have cancer. However, because there is a 2% false-positive rate for the test, 20 of these young males would have a false-positive result (2% of 999 = 19.98). The one male with cancer among the 1,000 young males would almost certainly yield a positive result as well. So, 21 of the young males would have a positive test result, but only one of the 21 would actually have cancer. Hence, the probability that any one of these young males who had a positive result actually has cancer would be only about 1 in 21, or 4.8%.

This same base-rate issue applies to police lineup tests. Specifically, the probability that a suspect is guilty given that the witness identified that suspect is a function of both the diagnostic value of the evidence and the base-rate probability that a lineup's suspect is guilty. Imagine one extreme jurisdiction (the "Bumbling Detectives PD") in which none of the lineups that police conduct include the guilty suspect (i.e., the target-present base rate is 0%). With a 0% base rate, even a miniscule false-positive rate yields only mistaken identifications and no accurate identifications. Now imagine the other extreme (the "Perfect Detectives PD"), a jurisdiction in which the suspect in a lineup is always the perpetrator (i.e., the target-present base rate is 100%). With a 100% base rate, even a high false-positive rate would yield no false positives on suspect identifications: Every ID of a suspect would be accurate.

When the base rate is 0%, the accuracy rate for identifications of the suspect is 0%, and when the base rate is 100%, the accuracy rate for identifications of suspect is 100%. Of course, real base rates for target-present lineups in police departments will lie somewhere between these two extremes. And, as one moves from the 0% base rate to the 100% base rate, the probability that the identified suspect is the perpetrator follows a Bayesian curve (not a straight line)—a prior-by-posterior probability curve.

Consider the prior-by-posterior curves that we created for the Wetmore et al. (2015) data as displayed in Figure 8. We used Bayes's theorem to calculate each point in these curves. Here, we show how three specific points on the moderate-confidence curve were calculated—one at the 30% base rate, one at the 50% base rate, and one at the 80% base rate.

The vertical axis in Figure 8 is the probability that the suspect is the perpetrator given that the witness identified the suspect from the lineup, which is what we are trying to estimate. We use the expression $p(SP|IDS)$ to represent the probability that the suspect is the perpetrator (SP) given an identification of the suspect (IDS). We use the expression $p(IDS|SP)$ to represent the probability of identification of the suspect (IDS) given that the suspect is the perpetrator (SP). In effect, $p(IDS|SP)$ is the hit rate. Likewise, $p(IDS|SNP)$ is the probability of identification of the suspect (IDS) given that the suspect is *not* the perpetrator (SNP). In effect, $p(IDS|SNP)$ is the false-alarm rate. The term $p(SP)$ is the target-present base rate (or prior probability that the suspect is the perpetrator). The term $p(SNP)$ is, in effect, the target-absent base rate, which is $1 - p(SP)$. We can then put the data into a version of Bayes's theorem as shown below.

$$p(SP \mid IDS) = \frac{p(IDS \mid SP) \times p(SP)}{\left(p(IDS \mid SP)\,p(SP)\right) + \left(p(IDS \mid SNP)\,p(SNP)\right)}$$

In the Wetmore et al. (2015) data, $p(IDS|SP)$ (i.e., the hit rate) for moderate-confidence witnesses was 72.3%, and $p(IDS|SNP)$ (i.e., the mistaken-identification rate) was 10.6%. These two values do not change as a function of the base rate. In effect, these two values constitute the diagnosticity of IDs by the moderate-confidence witnesses. Using the Bayesian expression in Equation 1, the probability that the suspect is the perpetrator given that the witness identified the suspect for the 50% base rate is:

$$p(SP \mid IDS) = \frac{.723 \times .50}{(.723 \times .50) + (.106 \times .50)} = .872$$

Suppose, however, that the base rate was 80%. The probability that the suspect is the perpetrator given that the witness identified the suspect for the 50% base rate is:

$$p(SP \mid IDS) = \frac{.723 \times .80}{(.723 \times .80) + (.106 \times .20)} = .965$$

And, if the base rate was 30%, the probability that the suspect is the perpetrator given that the witness identified the suspect is:

$$p(SP \mid IDS) = \frac{.723 \times .30}{(.723 \times .30) + (.106 \times .70)} = .745$$

Each of these three points on the moderate-confidence curve can be observed in Figure 8.

The degree to which base-rate changes (e.g., from 30% to 80%) moderate the probability that an identified suspect is guilty depends on the diagnosticity of the witness. As diagnosticity increases, the effect of the base rate diminishes. For example, for moderate-confidence witnesses in the Wetmore et al. data, moving from a 30% base rate to an 80% base rate changed the probability that the suspect was the perpetrator from 74.5% to 96.5%, a change of over

20 percentage points. But for high-confidence witnesses, moving from the 30% base rate to the 80% base rate changed the probability that the suspect was the perpetrator from 87.7% to 98.5%, a change of less than 11 percentage points. And for low-confidence witnesses, moving from the 30% base rate to the 80% base rate changed the probability that the suspect was the perpetrator from 62.8% to 94.0%, a change of over 30 percentage points.

Another observation about base rates of note here is that the confidence of the witness makes more difference to our ability to trust the identification when the base rate is in the lower ranges than when the base rate is in the upper ranges. Using the Wetmore et al. (2015) data, for example, when the base rate is 35%, the probability that an identified suspect is guilty for low-confidence witnesses is 22% lower than it is for high-confidence witnesses. When the base rate is 90%, however, the probability that an identified suspect is guilty for low-confidence witnesses is only 2% lower than it is for high-confidence witnesses. This means that when jurisdictions have lineups with relatively low target-present base rates, the importance of eyewitness confidence is even greater than when their lineups' base rates are higher.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Notes

1. Note that in these studies and in most of the ones we consider later, there are more IDs made with high than low confidence, so suspect-ID accuracy scores for low-confidence IDs tend to be more variable than for high-confidence IDs.
2. Only 5% of witnesses made IDs at the highest level (a rating of 7) of confidence, which makes the sample size unstable for isolating this one level of confidence. Hence, we combined confidence levels 6 and 7.

References

American Psychological Association. (2014). *Commonwealth v. Gomes and Commonwealth v. Johnson.* Retrieved from http://www.apa.org/about/offices/ogc/amicus/gomes-johnson.aspx

Behrman, B.W., & Davey, S.L. (2001). Eyewitness identification in actual criminal cases: An archival analysis. *Law and Human Behavior, 25*, 475–491.

Behrman, B. W., & Richards, R. E. (2005). Suspect/foil identification in actual crimes and in the laboratory: A reality monitoring analysis. *Law and Human Behavior, 29*, 279–301. doi:10.1007/s10979-005-3617-y

Berard, P. (2014). Eyewitness testimony at heart of court cases. *New England Psychologist.* Retrieved from http://www.nepsy.com/articles/leading-stories/eyewitness-testimony-at-heart-of-court-cases/

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. doi:10.1037/0021-9010.72.4.691

Bradfield, A., & McQuiston, D. E. (2004). When does evidence of eyewitness confidence inflation affect judgments in a criminal trial? *Law and Human Behavior, 28*, 369–387. doi:10.1037/1076-898x.8.1.44

Brady v. Maryland (1983). 373 U.S. 83.

Brainerd, C. J., & Reyna, V. F. (2002). Recollection rejection: How children edit their false memories. *Developmental Psychology, 38*, 156–172. doi:10.1037//0012-1649.38.1.156

Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory.* New York, NY: Oxford University Press.

Brewer, N., Keast, A., & Rishworth, A. (2002). The confidence-accuracy relationship in eyewitness identification: The effects of reflection and disconfirmation on correlation and calibration. *Journal of Experimental Psychology: Applied, 8*, 44–56. doi:10.1037/1076-898x.8.1.44

Brewer, N., & Palmer, M. A. (2010). Eyewitness identification tests. *Legal and Criminological Psychology, 15*, 77–96. doi:10.1348/135532509x414765

Brewer, N., Weber, N., Wootton, D., & Lindsay, D. (2012). Identifying the bad guy in a lineup using confidence judgments under deadline pressure. *Psychological Science, 23*, 1208–1214. doi:10.1177/0956797612441217

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relation in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied, 12*, 11–30. doi:10.1037/1076-898x.12.1.11

Brewer, N., & Wells, G. (2011). Eyewitness identification. *Current Directions in Psychological Science, 20*, 24–27. doi:10.1177/0963721410389169

Brodes v. State, 614 SE 2d 766 (2005).

Brown, E., Deffenbacher, K., & Sturgill, W. (1977). Memory for faces and the circumstances of the encounter. *Journal of Applied Psychology, 62*, 311–318. doi:10.1037/0021-9010.62.3.311

Buratti, S., & Allwood, C. M. (2012). Improved realism of confidence for an episodic memory event. *Judgment and Decision Making, 7*, 590–601.

Carlson, C. A., & Carlson, M. A. (2014). An evaluation of perpetrator distinctiveness, weapon presence, and lineup presentation using ROC analysis. *Journal of Applied Research in Memory and Cognition, 3*, 45–53. doi:10.1016/j.jarmac.2014.03.004

Carlson, C. A., Dias, J. L., Weatherford, D. R., & Carlson, M. A. (in press). An investigation of the weapon focus effect and the confidence–accuracy relationship for eyewitness identification. *Journal of Applied Research in Memory and Cognition.* doi:10.1016/j.jarmac.2016.04.001

Charman, S. D., & Wells, G. L. (2007). Applied lineup theory. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of eyewitness psychology: Memory for people* (pp. 219–254). Mahwah, NJ: Lawrence Erlbaum.

Charman, S. D., Wells, G. L., & Joy, S. W. (2011). The dud effect: Highly dissimilar fillers increase confidence in lineup identifications. *Law and Human Behavior, 35,* 479–500. doi:10.1007/s10979-010-9261-1

Charman, S. D., & Wells, G. L. (2012). The moderating effect of ecphoric experience on post-identification feedback: A critical test of the cues-based inference conceptualization. *Applied Cognitive Psychology, 26,* 243–250. doi:10.1002/acp.1815

Clark, S. E., & Tunnicliff, J. L. (2001). Selecting lineup foils in eyewitness identification experiments: Experimental control and real-world simulation. *Law and Human Behavior, 25,* 199–216. doi:10.1023/a:1010753809988

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

Clifford, B. R., & Scott, J. (1978). Individual and situational factors in eyewitness testimony. *Journal of Applied Psychology, 63,* 352–359. doi:10.1037/0021-9010.63.3.352

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

Colloff, M. F., Wade, K. A., & Strange, D. (2016). Unfair lineups make witnesses more likely to confuse innocent and guilty suspects. *Psychological Science, 27,* 1227–1239. doi:10.1177/0956797616655789

Davis, J. P., & Valentine, T. (2009). CCTV on trial: Matching video images with the defendant in the dock. *Applied Cognitive Psychology, 23,* 482–505. doi:10.1002/acp.1490

Deffenbacher, K. A. (1980). Eyewitness accuracy and confidence: Can we infer anything about their relationship? *Law and Human Behavior, 4,* 243–260. doi:10.1007/bf01040617

Deffenbacher, K. A., Brown, E. L., & Sturgill, W. (1978). Some predictors of eyewitness memory accuracy. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 219–226). New York, NY: Academic Press.

Dobolyi, D. G., & Dodson, C. S. (2013). Eyewitness confidence in simultaneous and sequential lineups: A criterion shift account for sequential mistaken identification overconfidence. *Journal of Experimental Psychology: Applied, 19,* 345–357. doi:10.1037/a0034596

Dodson, C. S., & Dobolyi, D. G. (2016). Confidence and eyewitness identifications: The cross-race effect, decision time and accuracy. *Applied Cognitive Psychology, 30,* 113–125. doi:10.1002/acp.3178

Dunning, D., & Perretta, S. (2002). Automaticity and eyewitness accuracy: A 10- to 12-second rule for distinguishing accurate from inaccurate positive identifications. *Journal of Applied Psychology, 87,* 951–962. doi:10.1037/0021-9010.87.5.951

Dysart, J. E., Lawson, V. Z., & Rainey, A. (2012). Blind lineup administration as a prophylactic against the post-identification feedback effect. *Law and Human Behavior, 36,* 312–319. doi:10.1037/h0093921

Fitzgerald, R. J., Price, H. L., Oriet, C., & Charman, S. (2013). The effect of suspect-filler similarity on eyewitness identification decisions: A meta-analysis. *Psychology, Public Policy, and Law, 19,* 151–164. doi:10.1037/a0030618

Garrett, B. (2011). *Convicting the innocent: Where criminal prosecutions go wrong.* Cambridge, MA: Harvard University Press.

Garrioch, L., & Brimacombe, C. A. E. (2001). Lineup administrators' expectations: Their impact on eyewitness confidence. *Law and Human Behavior, 25,* 299–315. doi:10.1023/a:1010750028643

Garry, M., Manning, C. G., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin & Review, 3,* 208–214. doi:10.3758/bf03212420

Gronlund, S. D., Carlson, C. A., Dailey, S. B., & Goodsell, C. A. (2009). Robustness of the sequential lineup advantage. *Journal of Experimental Psychology: Applied, 15,* 140–152. doi:10.1037/a0015082

Gronlund, S. D., Carlson, C. A., Neuschatz, J. S., Goodsell, C. A., Wetmore, S. A., Wooten, A., & Graham, M. (2012). Showups versus lineups: An evaluation using ROC analysis. *Journal of Applied Research in Memory and Cognition, 1,* 221–228. doi:10.1016/j.jarmac.2012.09.003

Horry, R., & Brewer, N. (2016). How target-lure similarity shapes confidence judgments in multiple-alternative decision tasks. *Journal of Experimental Psychology: General, 145,* 1615–1634. doi.10.1037/xge0000227

Horry, R., Palmer, M., & Brewer, N. (2012). Backloading in the sequential lineup prevents within-lineup criterion shifts that undermine eyewitness identification performance. *Journal of Experimental Psychology: Applied, 18,* 346–360. doi:10.1037/a0029779

Innocence Project. (2013). Brief of amicus curiae the Innocence Project in support of Thomas and Raymond Highers motion in limine for an eyewitness-identification expert. The People of the State of Michigan vs. Raymond Highers & Thomas Highers (Case No. 87-6345).

Innocence Project. (2016). DNA Exonerations in the United States. Retrieved January 17, 2017. http://www.innocenceproject.org/dna-exonerations-in-the-united-states/

Jones, E. E., Williams, K. D., & Brewer, N. (2008). "I had a confidence epiphany!": Obstacles to combating post-identification confidence inflation. *Law and Human Behavior, 32,* 164–176. doi:10.1007/s10979-007-9101-0

Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence-accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1304–1316. doi:10.1037/0278-7393.22.5.1304

Kassin, S. M. (1998). Eyewitness identification procedures: The fifth rule. *Law and Human Behavior, 23,* 649–654. doi:10.1023/a:1025702722645

Keast, A., Brewer, N., & Wells, G. L. (2007). Children's metacognitive judgments in an eyewitness identification task. *Journal of Experimental Child Psychology, 97,* 286–314. doi:10.1016/j.jecp.2007.01.007

Klobuchar, A., Steblay, N. K., & Caligiuri, H. L. (2006). Improving eyewitness identification: Hennepin County's blind sequential lineup pilot project. *Cardozo Public Law, Policy, and Ethics Journal, 2,* 381–414.

Lampinen, J. M., Erickson, W. B., Moore, K. N., & Hittson, A. (2014). Effects of distance on face recognition: Implications for eyewitness identification. *Psychonomic Bulletin & Review, 21,* 1489–1494. doi:10.3758/s13423-014-0641-2

Lampinen, J. M., Neuschatz, J. S., & Cling, A. D. (2012). *The psychology of eyewitness identification.* New York, NY: Psychology Press.

Leach, A., Cutler, B. L., & Van Wallendael, L. (2009). Lineups and eyewitness identification. *Annual Review of Law and*

Social Science, 5, 157–178. doi:10.1146/annurev.lawsocsci.093008.131529

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. doi:10.1037//0021-9010.63.3.345

Lindsay, D. S. (2014). Memory source monitoring applied. In T. Perfect & D. S. Lindsay (Eds.), *The SAGE handbook of applied memory* (pp. 59–75). London, England: Sage.

Lindsay, D. S., Nilsen, E., & Read, J. D. (2000). Witnessing-condition heterogeneity and witnesses' versus investigators' confidence in the accuracy of witnesses' identification decisions. *Law and Human Behavior, 24*, 685–697. doi:10.1023/a:1005504320565

Lindsay, D. S., Read, J. D., & Sharma, K. (1998). Accuracy and confidence in person identification: The relationship is strong when witnessing conditions vary widely. *Psychological Science, 9*, 215–218. doi:10.1111/1467-9280.00041

Lindsay, R. C. L. (1986). Confidence and accuracy in eyewitness identification from lineups. *Law and Human Behavior, 10*, 229–239. doi:10.1007/bf01046212

Lindsay, R. C. L., Martin, R., & Webber, L. (1994). Default values in eyewitness descriptions: A problem for the match-to-description lineup foil selection strategy. *Law and Human Behavior, 18*, 527–541. doi:10.1007/bf01499172

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. doi:10.1007/bf01040622

Luus, C. A. E., & Wells, G. L. (1991). Eyewitness identification and the selection of distracters for lineups. *Law and Human Behavior, 15*, 43–57. doi:10.1007/bf01044829

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. doi:10.1037/0021-9010.79.5.714

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago Face Database: A free stimulus set of faces and norming data. *Behavior Research Methods, 47*, 1122–1135.

Manson v. Braithwaite, 432 U.S. 98, 1977.

Massachusetts Court System. (2015). Criminal model jury instructions, 9160-Identification. Retrieved from http://www.mass.gov/courts/docs/courts-and-judges/courts/district-court/jury-instructions-criminal/6000-9999/9160-defenses-identification.pdf

Mickes, L. (2015). Receiver operating characteristic analysis and confidence-accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory and Cognition, 4*, 93–102. doi:10.1016/j.jarmac.2015.01.003

Mickes, L., Flowe, H. D., & Wixted, J. T. (2012). Receiver operating characteristic analysis of eyewitness memory: Comparing the diagnostic accuracy of simultaneous vs. sequential lineups. *Journal of Experimental Psychology: Applied, 18*, 361–376. doi:10.1037/a0030609

Mickes, L., Hwe, V., Wais, P. E., & Wixted, J. T. (2011). Strong memories are hard to scale. *Journal of Experimental Psychology: General, 140*, 239–257. doi:10.1037/a0023007

Münsterberg, H. (1908). *On the witness stand: Essays on psychology and crime.* Garden City, NY: Clark, Boardman.

National Institute of Justice. (1999). *Eyewitness evidence: A guide for law enforcement.* Washington, DC: Office of Justice Programs, U.S. Department of Justice.

National Research Council. (2014). *Identifying the culprit: Assessing eyewitness identification*. Washington, DC: The National Academies Press.

Neal, T. M. S., Christiansen, A., Bornstein, B. H., & Robicheaux, T. R. (2012). The effects of mock jurors' beliefs about eyewitness performance on trial judgments. *Psychology, Crime & Law, 18*, 49–64. doi:10.1080/1068316x.2011.587815

New Jersey Courts. (2012a). Identification - in court and out of court identifications [New Jersey criminal model jury charges]. Retrieved from http://www.judiciary.state.nj.us/pressrel/2012/jury_instruction.pdf

New Jersey Courts. (2012b, July 19). *Supreme Court releases eyewitness identification criteria for criminal cases* [Press release]. Retrieved from http://www.judiciary.state.nj.us/pressrel/2012/pr120719a.htm

Palmer, M., Brewer, N., Weber, N., & Nagesh, A. (2013). The confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention. *Journal of Experimental Psychology: Applied, 19*, 55–71. doi:10.1037/a0031602

Penrod, S., & Cutler, B. (1995). Witness confidence and witness accuracy: Assessing their forensic relation. *Psychology, Public Policy, and Law, 1*, 817–845. doi:10.1037/1076-8971.1.4.817

Police Executive Research Forum. (2013). A national survey of eyewitness identification procedures in law enforcement agencies. Retrieved from http://www.policeforum.org/assets/docs/Free_Online_Documents/Eyewitness_Identification/a%20national%20survey%20of%20eyewitness%20identification%20procedures%20in%20law%20enforcement%20agencies%202013.pdf

Quinlivan, D. S., Neuschatz, D. S., Wells, G. L., Cutler, B. L., McClung, J. E., & Harker, D. (2012). Do pre-admonition suggestions moderate the effect of the unbiased-lineup instructions? *Legal and Criminological Psychology, 17*, 165–176. doi:10.1348/135532510x533554

Read, J. D., Lindsay, D. S., & Nichols, T. (1998). The relation between confidence and accuracy in eyewitness identification studies: Is the conclusion changing? In C. P. Thomson, D. Bruce, J. D. Read, D. Hermann, D. Payne, & M. P. Toglia (Eds.), *Eyewitness memory: Theoretical and applied perspectives* (pp. 107–130). Mahwah, NJ: Lawrence Erlbaum.

Read, J. D., Tollestrup, P., Hammersley, R., McFadzen, E., & Christensen, A. (1990). The unconscious transference effect: Are innocent bystanders ever misidentified? *Applied Cognitive Psychology, 4*, 3–31. doi:10.1002/acp.2350040103

Read, J. D., Yuille, J. C., & Tollestrup, P. (1992). Recollections of a robbery: Effects of alcohol and arousal upon recall and person identification. *Law and Human Behavior, 16*, 425–446. doi:10.1007/bf02352268

Reinitz, M. T., Seguin, J. A., Peria, W., & Loftus, G. R. (2012). Confidence-accuracy relations for faces and scenes: Roles of features and familiarity. *Psychonomic Bulletin & Review, 19*, 1085–1093. doi:10.3758/s13423-012-0308-9

Report of the Special Master, State v. Henderson, No. A-8-08. (2011, June 18). Retrieved from http://www.judiciary

.state.nj.us/pressrel/HENDERSON%20FINAL%20BRIEF%20.PDF%20%2800621142%29.PDF

Rodriguez, D. N., & Berry, M. A. (2014). The effect of line-up administrator blindness on recording of eyewitness identification decisions. *Legal and Criminological Psychology, 19*, 69–79. doi:10.1111/j.2044-8333.2012.02058.x

Roediger, H. L., & DeSoto, K. A. (2015). Understanding the relation between confidence and accuracy in reports from memory. In D. S. Lindsay, C. M. Kelley, A. P. Yonelinas, & H. L. Roediger (Eds.), *Remembering: Attributions, processes, and control in human memory: Papers in honor of Larry L. Jacoby* (pp. 347–367). New York, NY: Psychology Press.

Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences, 3*, 377–386. doi:10.1017/s0140525x00075506

Rosnow, R. L., Rosenthal, R., & Rubin, D. B. (2000). Contrasts and correlations in effect-size estimation. *Psychological Science, 11*, 446–453. doi:10.1111/1467-9280.00287

Ross, J. E. (2006). The entrenched position of plea bargaining in United States legal practice. *The American Journal of Comparative Law, 54*, 717–732. doi:10.2307/20454559

Sauer, J., Brewer, N., & Weber, N. (2008). Multiple confidence estimates as indices of eyewitness memory. *Journal of Experimental Psychology: General, 137*, 528–554. doi:10.1037/a0012712

Sauer, J., Brewer, N., & Weber, N. (2012). Using confidence ratings to identify a target among foils. *Journal of Applied Research in Memory and Cognition, 1*, 80–88. doi:10.1016/j.jarmac.2012.03.003

Sauer, J., Brewer, N., & Wells, G. L. (2008). Is there a magical time boundary for diagnosing eyewitness identification accuracy in sequential line-ups? *Legal and Criminological Psychology, 13*, 123–135. doi:10.1348/135532506x159203

Sauer, J., Brewer, N., Zweck, T., & Weber, N. (2010). The effect of retention interval on the confidence-accuracy relationship for eyewitness identification. *Law and Human Behavior, 34*, 337–347. doi:10.1007/s10979-009-9192-x

Sauer, J., Weber, N., & Brewer, N. (2012). Using ecphoric confidence ratings to discriminate seen from unseen faces: The effects of retention interval and distinctiveness. *Psychonomic Bulletin & Review, 19*, 490–498. doi:10.3758/s13423-012-0239-5

Sauerland, M., & Sporer, S. L. (2009). Fast and confident: Postdicting eyewitness identification accuracy. *Journal of Experimental Psychology: Applied, 15*, 46–62. doi:10.1037/a0014560

Smalarz, L., & Wells, G. L. (2015). Contamination of eyewitness self-reports and the mistaken-identification problem. *Current Directions in Psychological Science, 24*, 120–124. doi:10.1177/0963721414554394

Smith, H. M. J., & Flowe, H. D. (2014). ROC analysis of the verbal overshadowing effect: Testing the effect of verbalization on memory sensitivity. *Applied Cognitive Psychology, 29*, 159–168. doi:10.1002/acp.3096

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology, 78*, 22–33. doi:10.1037/0021-9010.78.1.22

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence–accuracy relation in eyewitness identification studies. *Psychological Bulletin, 118*, 315–327. doi:10.1037/0033-2909.118.3.315

State v. Guilbert. Supreme Court of Connecticut, 306 Conn. 218 (2012).

State v. Lawson (2012). 352 Or. 724, 291 P.3d 673.

State v. Mitchell, 275 P.3d 905 (Kan. 2012).

Steblay, N. M., Wells, G. L., & Douglass, A. L. (2014). The eyewitness post identification feedback effect 15 years later: Theoretical and policy implications. *Psychology, Public Policy, and Law, 20*, 1–18. doi:10.1037/law0000001

Sučić, I., Tokić, D., & Ivešić, M. (2015). Field study of response accuracy and decision confidence with regard to lineup composition and lineup presentation. *Psychology, Crime & Law, 21*, 798–819. doi:10.1080/1068316x.2015.1054383

Thompson, J. (2000, June 18). I was certain, but I was wrong. *The New York Times.* Retrieved from http://www.nytimes.com/2000/06/18/opinion/i-was-certain-but-i-was-wrong.html

Thompson-Cannino, J. T., Cotton, R., & Torneo, E. (2009). *Picking Cotton: Our memoir of justice and redemption.* New York, NY: St. Martin's Press.

Tollestrup, P. A., Turtle, J. W., & Yuille, J. C. (1994). Actual 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. 144–162). New York, NY: Cambridge University Press.

U.S. Department of Justice. (1999). *Eyewitness evidence: A guide for law enforcement.* Washington, DC: Office of Justice Programs.

Valentine, T., & Davis, J. P. (2015). *Forensic facial identification: Theory and practice of identification from eyewitnesses, composites and CCTV.* West Sussex, UK: John Wiley.

Weber, N., & Brewer, N. (2004). Confidence-accuracy calibration in absolute and relative face recognition judgments. *Journal of Experimental Psychology: Applied, 10*, 156–172. doi:10.1037/1076-898x.10.3.156

Wells, G. L. (1978). Applied eyewitness testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology, 36*, 1546–1557. doi:10.1037/0022-3514.36.12.1546

Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology, 14*, 89–103. doi:10.1111/j.1559-1816.1984.tb02223.x

Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist, 48*, 553–571. doi:10.1037/0003-066x.48.5.553

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. doi:10.1037/0021-9010.83.3.360

Wells, G. L., & Bradfield, A. L. (1999). Distortions in eyewitnesses' recollections: Can the post-identification feedback effect be moderated? *Psychological Science, 10*, 138–144. doi:10.1111/1467-9280.00121

Wells, G. L., Ferguson, T. J., & Lindsay, R. C. L. (1981). The tractability of eyewitness confidence and its implication for triers of fact. *Journal of Applied Psychology, 66*, 688–696. doi:10.1037/0021-9010.66.6.688

Wells, G. L., Leippe, M. R., & Ostrom, T. M. (1979). Guidelines for empirically assessing the fairness of a lineup. *Law and Human Behavior, 3*, 285–293. doi:10.1007/bf01039807

Wells, G. L., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness non-identifications. *Psychological Bulletin, 88*, 776–784. doi:10.1037/0033-2909.88.3.776

Wells, G. L., & Luus, E. (1990). Police lineups as experiments: Social methodology as a framework for properly-conducted lineups. *Personality and Social Psychology Bulletin, 16*, 106–117. doi:10.1177/0146167290161008

Wells, G. L., 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., Rydell, S. M., & Seelau, E. P. (1993). On the selection of distractors for eyewitness lineups. *Journal of Applied Psychology, 78*, 835–844. doi:10.1037//0021-9010.78.5.835

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, 23*, 603–647. doi:10.1023/a:1025750605807

Wells, G. L., & Turtle, J. W. (1986). Eyewitness identification: The importance of lineup models. *Psychological Bulletin, 99*, 320–329. doi:10.1037/0033-2909.99.3.320

Wells, G. L., Yang, Y., & Smalarz, L. (2015). Eyewitness identification: Bayesian information gain, base-rate effect equivalency curves, and reasonable suspicion. *Law and Human Behavior, 39*, 99–122. doi:10.1037/lhb0000125

Wetmore, S. A., Neuschatz, J. S., Gronlund, S. D., Wooten, A., Goodsell, C. A., & Carlson, C. A. (2015). Effect of retention interval on showup and lineup performance. *Journal of Applied Research in Memory and Cognition, 4*, 8–14. doi:10.1016/j.jarmac.2014.07.003

Wilson, J. P., Hugenberg, K., & Bernstein, M. J. (2013). The cross-race effect and eyewitness identification: How to improve recognition and reduce decision errors in eyewitness situations. *Social Issues and Policy Review, 7*, 83–113. doi:10.1111/j.1751-2409.2012.01044.x

Wise, R. A., Safer, M. A., & Maro, C. M. (2011). What U.S. law enforcement officers know and believe about eyewitness factors, eyewitness interviews, and identification procedures. *Applied Cognitive Psychology, 25*, 488–500. doi:10.1002/acp.1717

Wixted, J. T., Mickes, L., Clark, S. E., Dunn, J. C., & Wells, W. (2016). Estimating the reliability of eyewitness identifications from police lineups. *Proceedings of the National Academy of Sciences, USA, 113*, 304–309. doi:10.1073/pnas.1516814112

Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., & Roediger, H. L. (2015). Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *American Psychologist, 70*, 515–526. doi:10.1037/a0039510

Wixted, J. T., Read, J. D., & Lindsay, D. S. (2016). The effect of retention interval on the eyewitness identification confidence-accuracy relationship. *Journal of Applied Research in Memory and Cognition, 5*, 192–203.

EXHIBIT 6

Northeastern Illinois Regional Crime Laboratory

STRmix Report

905 E. Orchard St., Mundelein, IL 60060
Phone: (847) 362-0676 Fax: (847) 362-0712

Board President
Steve Husak

Executive Director
Philip T. Kinsey, Ph.D.

ANAB ANSI National Accreditation Board ACCREDITED ISO/IEC 17025 FORENSIC TESTING LABORATORY

Chief Laz Perez
North Chicago Police Department
1850 Lewis Avenue
North Chicago, IL 60064

Subject: Homicide Investigation
Agency Case #: 00-002716
Case Officer: SAL CECALA
Submission Date: 12/10/2013, 04/08/2004, 03/23/2004, 01/17/2014

Laboratory Case #: 00-483
Laboratory Report #: 20
Report Date: 05/18/2022

Case Names: DELWIN NMI FOXWORTH SR
MARVIN T WILLFORD

The following evidence was submitted/retained in a sealed condition:

ITEM 01 (1SC)	Exhibit 01	Extracted DNA from wood (stain B, report 18)
	Exhibit 01	Extracted DNA from wood (stain B, second extraction, report 19)
	Exhibit 01	Extracted DNA from wood (stain D, report 18)
	Exhibit 01	Extracted DNA from wood (stain G, second extraction, report 19)
ITEM 02 (16SC)	Exhibit 01	Extracted DNA from gas can (stain C, report 19)
	Exhibit 01	Extracted DNA from gas can (stain E, report 19)
ITEM 15 (7SC)	Exhibit 01	Extracted DNA from right boot (stain D, report 13)
ITEM 26 (4DB)	Exhibit 01	Extracted DNA from duct tape (stain A, report 13)
ITEM 37 (13DB)	Exhibit 01	Extracted DNA from the known standard of "Amarin Willford"
ITEM 38 (11DB)	Exhibit 01	Extracted DNA from the known standard of "Delwin L. Foxworth"

DNA RESULTS

DNA profiling of the above DNA extracts was conducted by PCR using the GlobalFiler amplification kit, containing 21 STR loci, one YSTR locus, one Y Indel and Amelogenin.

Partial DNA profiles, each originating from multiple contributors, were obtained from wood (stain G, second extraction), wood (stain D), gas can (stain C), gas can (stain E), right boot (stain D) and duct tape (stain A).

DNA profiles, each originating from multiple contributors, were obtained from wood (stain B,

Lab Case 00-483
Lab Report # 20
Analyst Maria A. Salazar, B.S.

second extraction) and wood (stain B).

DNA profiles were obtained from the known standards of "Amarin Willford" and "Delwin L. Foxworth".

Please note that two manipulation blanks did not perform as expected. The D5S818 locus for duct tape (stain A), and the SE33 and D10S1248 loci for wood (stain B) and wood (stain D) will not be used for comparison analysis.

STRmix RESULTS

The above profiles were evaluated using STRmix™, a probabilistic genotyping software application.

Item # 26.01A Description: duct tape (stain A)

Assumed number of contributors: 2
Assumed contributor(s): N/A

Person of Interest	Likelihood Ratio (LR)	Level of Support
"Amarin Willford"	0	Exclusion
"Delwin L. Foxworth"	2.63 Billion	Very Strong Support for Inclusion

"Amarin Willford" is excluded as a possible contributor to the partial DNA profile obtained from duct tape (stain A) (20 of 21 loci were used in this calculation).

The partial mixture of DNA obtained from duct tape (stain A) is 2.63 Billion times more likely if it originated from "Delwin L. Foxworth" and one unknown contributor than if it originated from two unknown contributors (20 of 21 loci were used in this calculation).

Item # 15.01D Description: right boot (stain D)

Assumed number of contributors: 3
Assumed contributor(s): N/A

Person of Interest	Likelihood Ratio (LR) and (1/LR)	Level of Support
"Amarin Willford"	109	Moderate Support for Exclusion
"Delwin L. Foxworth"	4.11 Million	Very Strong Support for Inclusion

The partial mixture of DNA obtained from right boot (stain D) is 109 times more likely if it originated from three unknown contributors than if it originated from "Amarin Willford" and two unknown contributors.

The partial mixture of DNA obtained from right boot (stain D) is 4.11 Million times more likely if it originated from "Delwin L. Foxworth" and two unknown contributors than if it originated from three unknown contributors.

Lab Case 00-483
Lab Report # 20
Analyst Maria A. Salazar, B.S.

Item # 01.01B Description: wood (stain B)

Assumed number of contributors: 3
Assumed contributor(s): N/A

Person of Interest	Likelihood Ratio (LR) and (1/LR)	Level of Support
"Amarin Willford"	358	Moderate Support for Exclusion
"Delwin L. Foxworth"	91.7 Sextillion	Very Strong Support for Inclusion

The mixture of DNA obtained from wood (stain B) is 358 times more likely if it originated from three unknown contributors than if it originated from "Amarin Willford" and two unknown contributors (19 of 21 loci were used in this calculation).

The mixture of DNA obtained from wood (stain B) is 91.7 Sextillion times more likely if it originated from "Delwin L. Foxworth" and two unknown contributors than if it originated from three unknown contributors (19 of 21 loci were used in this calculation).

Item # 01.01D Description: wood (stain D)

Assumed number of contributors: 3
Assumed contributor(s): N/A

Person of Interest	Likelihood Ratio (LR) and (1/LR)	Level of Support
"Amarin Willford"	8.56	Limited Support for Exclusion
"Delwin L. Foxworth"	368 Quintillion	Very Strong Support for Inclusion

The partial mixture of DNA obtained from wood (stain D) is 8.56 times more likely if it originated from three unknown contributors than if it originated from "Amarin Willford" and two unknown contributors (19 of 21 loci were used in this calculation).

The partial mixture of DNA obtained from wood (stain D) is 368 Quintillion times more likely if it originated from "Delwin L. Foxworth" and two unknown contributors than if it originated from three unknown contributors (19 of 21 loci were used in this calculation).

Item # 02.01C Description: gas can (stain C)

Assumed number of contributors: 4
Assumed contributor(s): N/A

Person of Interest	Likelihood Ratio (1/LR)	Level of Support
"Amarin Willford"	12.7	Limited Support for Exclusion
"Delwin L. Foxworth"	195	Moderate Support for Exclusion

The partial mixture of DNA obtained from gas can (stain C) is 12.7 times more likely if it originated from four unknown contributors than if it originated from "Amarin Willford" and three unknown contributors.

Lab Case 00-483
Lab Report # 20
Analyst Maria A. Salazar, B.S.

The partial mixture of DNA obtained from gas can (stain C) is 195 times more likely if it originated from four unknown contributors than if it originated from "Delwin L. Foxworth" and three unknown contributors.

Item # 02.01E Description: gas can (stain E)

Assumed number of contributors: 3
Assumed contributor(s): N/A

Person of Interest	Likelihood Ratio (LR) and (1/LR)	Level of Support
"Amarin Willford"	13.6 Thousand	Strong Support for Exclusion
"Delwin L. Foxworth"	0	Exclusion

The partial mixture of DNA obtained from gas can (stain E) is 13.6 Thousand times more likely if it originated from three unknown contributors than if it originated from "Amarin Willford" and two unknown contributors.

"Delwin L. Foxworth" is excluded as a possible contributor to the partial DNA profile obtained from gas can (stain E).

Item #01.01BE2 Description: wood (stain B, second extraction)

Assumed number of contributors: 3
Assumed contributor(s): N/A

Person of Interest	Likelihood Ratio (LR) and (1/LR)	Level of Support
"Amarin Willford"	280	Moderate Support for Exclusion
"Delwin L. Foxworth"	2.72 Septillion	Very Strong Support for Inclusion

The mixture of DNA obtained from wood (stain B, second extraction) is 280 times more likely if it originated from three unknown contributors than if it originated from "Amarin Willford" and two unknown contributors.

The mixture of DNA obtained from wood (stain B, second extraction) is 2.72 Septillion times more likely if it originated from "Delwin L. Foxworth" and two unknown contributors than if it originated from three unknown contributors.

Item #01.01GE2 Description: wood (stain G, second extraction)

Assumed number of contributors: 3
Assumed contributor(s): N/A

Person of Interest	Likelihood Ratio (LR) and (1/LR)	Level of Support
"Amarin Willford"	1.70 Thousand	Moderate Support for Exclusion
"Delwin L. Foxworth"	22.3 Septillion	Very Strong Support for Inclusion

Lab Case 00-483
Lab Report # 20
Analyst Maria A. Salazar, B.S.

The partial mixture of DNA obtained from wood (stain G, second extraction) is 1.70 Thousand times more likely if it originated from three unknown contributors than if it originated from "Amarin Willford" and two unknown contributors.

The partial mixture of DNA obtained from wood (stain G, second extraction) is 22.3 Septillion times more likely if it originated from "Delwin L. Foxworth" and two unknown contributors than if it originated from three unknown contributors.

REMARKS

The propositions were formed from the information available to the undersigned at the time of analysis. If this information changes or other propositions should be considered, the analyst is able to undertake them if instructed with sufficient time.

The verbal scales listed in the tables below were implemented for use with STRmix™ results only. Verbal scales are designed to assist in conveying the weight of likelihood ratios. Equal (or nearly equal) support for both propositions results in a likelihood ratio of 1, which is qualified as Uninformative. As likelihood ratios increase in magnitude, the scale reflects stronger degrees of support.

Likelihood Ratio (LR) for H_1 Support and 1/LR for H_2 Support	Verbal Qualifier
1	Uninformative
2 – 99	Limited Support
100 – 9,999	Moderate Support
10,000 – 999,999	Strong Support
≥ 1,000,000	Very Strong Support

DNA evidence will be maintained at the laboratory should further analysis be required.

Should further analysis be required, please contact this examiner.

The results portion of this report contains scientific judgments and interpretations rendered by the individual whose signature appears on the report.

Please pick up all appropriate exhibits at your earliest convenience.

Reviewer
Sarah G. Ozanick, Ph.D.

Forensic Scientist
Maria A. Salazar, B.S.

EXHIBIT 7

Independent Forensics
DNA TESTING & TECHNOLOGIES

500 Waters Edge, Suite 210
Lombard IL 60148-7000
T: 708-234-1200 F: 708-978-5115
info@ifi-test.com www.ifi-test.com

January 23, 2024

David B. Owens
Staff Attorney
The Exoneration Project

e david@exonerationproject.org
c (312) 590-5449

Re: IL v Marvin T Williford — Court Docket 00 CF 1920, Lake County

Re: Northeastern Illinois Regional Crime Laboratory Case# 00-483

Documents Reviewed

-Northern Illinois Police Crime Lab, Laboratory Case Number 00-483, Report 1, Report Date 2/7/00

Latent print examinations

-Illinois State Police, Division of Forensic Services, Chicago, Laboratory Case# C00-007208, Dated February 24, 2000

chemical analysis for accelerant (gasoline)

-Northern Illinois Police Crime Lab, Laboratory Case Number 00-483, Report 4, Report Date 3/30/00

chemical analysis for controlled substances

-Northern Illinois Police Crime Lab, Laboratory Case Number 00-483, Report 5 Report Date 7/20/00

Item 01, exhibit 01 one piece of wood with notches at ends, retained

-Northern Illinois Police Crime Lab, Laboratory Case Number 00-483, Report 6, Report Date 7/19/2003

Latent print examination

-Northern Illinois Police Crime Lab, Laboratory Case Number 00-483, Report 7, Report Date 8/19/2003

Latent print examination

-Northern Illinois Police Crime Lab, Laboratory Case Number 00-483, Report 8, Report Date 12/20/2003

Comparison of latent ridge impressions with latent lifts, D.L. Foxworth

-Northern Illinois Police Crime Lab, Laboratory Case Number 00-483, Report 9, Report Date 1/30/04

chemical detection of blood; trace evidence

Notes for Report Number 9

-Northern Illinois Police Crime Lab, Laboratory Case Number 00-483, Report 10, Report Date 4/5/04

chemical detection of blood – Exhibit 01, stain C from wood

retained items of evidence 28-01 (sweatpants), 29-01 (sweatshirt), 38-01 (standard from D.L. Foxworth)

Allele summary charts, Laboratory Case Number 00-483, Notes for Report 10

-Northern Illinois Police Crime Lab, Laboratory Case Number 00-483, Report 11, Report Date 4/20/04

item/exhibit 01-01 (wood) retained for potential DNA; 36-01 (standard J. Adams), 37-01 (standard Amarin Willford), retained for comparison purposes

DNA analysis of item/exhibit 01-01 (wood): 36-01 (J. Adams) and 37-01 (A. Willford) excluded as source of the DNA. 38-01 (D.L. Foxworth) is not excluded as minor contributor.

Comparison of 36-01 (standard J. Adams) and 37-01 (standard A. Willford) to 01-01 (wood) stain C and to 29-01 (sweatshirt, stain A and stain B) excludes both J. Adams and A. Willford.

Allele summary charts, Laboratory Case Number 00-483, Notes for Report 11

-Independent Forensics, IFI Lab Case# NL-27730, Test Report, Dated January 22, 2014

Exhibit # 1, red gas can

Exhibit #2, 4 foot piece of wood

–Independent Forensics, IFI Lab Case# NL-27730, Supplemental Test Report, Dated April 3, 2014

Exhibit #1, handle of red gas can, DNA profile, comparison with reference standard

–Northeastern Illinois Regional Crime Laboratory, DNA Report, Laboratory Case# 00-483,Laboratory Report # 13, Report Date 05/13/2014

Item 01, exhibit 01 (stains from wood)
Item 15, pair of boots
Item 21, duct tape
Item 22, two pieces of duct tape
Item 26, wad of duct tape
Item 29, exhibit 01, stain A, sweatshirt (report 10)
Updated autosomal (CODIS) DNA-PCR-CE profiling kit (Identifiler)
Sex chromosome (Y-STR profiling, Y-Filer)
Re-extraction from retained stains and standards
J. Adams and A. Willford excluded as contributors to right boot (autosomal and Y-STR analysis)
01.01, second extraction from wood, Y-STR analysis, J. Adams, A, Willford, D.L. Foxworth excluded as contributors to

–LABSYS120127-00483 Rpt 13 DNA Packet

laboratory work sheets

–Northeastern Illinois Regional Crime Laboratory, CODIS Report, Laboratory Case# 00-483,Laboratory Report # 14, Report Date 06/05/2014

Case to case 'hits': 01.01 (wood, second extraction) to
(1) Lake County Sheriff's Case#92-55313, and to
(2) "Pooled Spermatozoa from Holly Staker Vaginal Swabs"

–Northeastern Illinois Regional Crime Laboratory, DNA Report, Laboratory Case# 00-483,Laboratory Report # 15, Report Date 06/16/2014

Contamination of questioned item 29-01, sweatshirt, stain A by laboratory personnel

–Northeastern Illinois Regional Crime Laboratory, Fingerprint Report, Laboratory Case# 00-483,Laboratory Report # 16, Report Date 10/8/2014

Processing for latent impressions

-Northeastern Illinois Regional Crime Laboratory, DNA Report,
Laboratory Case# 00-483, Laboratory Report # 17, Report Date 12/02/2014

Item 04, cordless phone
Item 11, broken cd (11.08), glass (11.02), glass (11.01), box (11.05), soap dispenser (11.06), diaper pack (11.07), crystal bowl (11.03),
Item 14, black wallet
item 17, table lamp
Item 19, one matchbook
Item 25, purple lighter
Item 39, wallet chain (01), clip-on earrings (02)
J. Adams and A, Willword excluded.
Laboratory backtracks on previous interpretation of 01.01, wood, first extraction

-LABSYS120130-00-483 Prt 17 DNA Packet 1

-Northeastern Illinois Regional Crime Laboratory, DNA Report,
Laboratory Case# 14-4780, Laboratory Report # 4, Report Date 04/02/2015

Submission of standards for comparison (many): all excluded from Lake County Sheriff's Case#92-5513, Pooled Spermatozoa from Holly Staker Vaginal Swabs, wood (01.01, second extraction), gas can handle (Independent Forensics 27730-80967-Q1).

LABSYS120134-00-483 Rpt 19 DNA Packet

-Northeastern Illinois Regional Crime Laboratory, Laboratory Report # 20
STRmix Report, Report Date 05/18/2022

Probabilistic mixture interpretation software analysis of previously obtained DNA profiles obtained from questioned items:
Item 01
~~Exhibit 01 – DNA extract from wood (stain B, report 18)~~
Exhibit 01 – DNA extract from wood (stain B, second extraction, report 19)
Exhibit 01 – DNA extracted from wood (stain D, report 18)
Exhibit 01 – DNA extracted from wood (stain G, second extraction, report 19)
Item 02
Exhibit 01 –DNA extract from gas can (stain C, report 19)
Exhibit 01 –DNA extract from gas can (stain E, report 19)
Item 15
Exhibit 01 –DNA extract from right boot (stain D, report 13)
Item 26
Exhibit 01 –DNA extract from duct tape (stain A, report 13)

Continued from previous page: Northeastern Illinois Regional Crime Laboratory, Laboratory Report # 20

Item 37

Exhibit 01 –DNA extract from known standard Amarin Willford

Item 38

Exhibit 01 –DNA extract from known standard Delwin Foxworth

Probabilistic mixture interpretation computer software analysis of previously obtained DNA profiles from 01.01 (wood) including stains B (both extractions, stain D and stain G (second extraction)

Also analyzed 02.01 (gas can) stain D and stain E.

Also analyzed 15.01 (stain from boot),

Also analyzed 26.01 (duct tape) stain A

THIS SPACE LEFT INTENTIONALLY BLANK

Report of Findings

This report is in supplement to and builds upon my prior work in this case, which remains valid (and subject to a clarification in my prior testimony about a small aspect of my prior report concerning DNA evidence on item of evidence, duct tape).

The forensic laboratory analysis and testing performed in this case encompasses several forensic disciplines including latent examination, chemical analysis and forensic DNA over the course of twenty-four (and counting) years.

While an overview and survey of the methods, types and results from the forensic DNA testing, data, and results are provided below, a clear and consistent conclusion from the twenty plus (20+) forensic DNA laboratory reports that detail the analysis of forty plus (40+) items of evidence can be readily summarized.

There is absolutely no forensic DNA evidence that identifies, links or ties the defendant, Mr. Williford, to any item of evidence in this case.

Given the sensitivity and specificity of modern forensic DNA methods and the results of the repeated efforts to obtain DNA profiling results from previously tested items as well as from newly submitted items, the lack of any link to Mr. Williford and the repeated exclusion of the defendant from the most probative items of evidence is itself probative and from a scientific perspective, speaks even more strongly than the results obtained from the original forensic DNA analysis pior to trial including the additional testing performed ca. 2014.

Given the range of samples tested, the various types of testing that have been performed and the repeated testing of the most probative items of evidence, it can also be stated that his DNA is not present on any of the evidence in this case.

The primary finding of this report, explained in detail below, is that the new 2020 forensic analysis, in conjunction with prior testing, provides probative evidence that Williford's DNA was not found at the crime scene. This conclusion is supported by fact that later testing modalities were far more sensitive and more discriminating that what was available, ca. 2000. This increase in specificity includes most recent software analysis performed by the Northeastern Regional Crime Laboratory.

Initial forensic analysis

The first laboratory reports in this case describe an effort to obtain and compare latent ridge impressions (fingerprints) recovered from the crime scene.

The review of this work is outside the purview of the author, but the conclusions reported for the examination of the latents lifts that were declared suitable for comparison (a quality standard for latent examination) are clear: Mr. Williford was excluded.

A short time later the first forensic biology results were reported. Here the presumptive identification of blood was described as were the first forensic DNA profiling results.

The laboratory, correctly and appropriately, excluded Mr. Williford from several highly probative items of evidence including item 01, exhibit 01, wood.

The laboratory would return to this item more than once.

It can be noted that the decedent victim was not excluded as a contributor to this item of evidence..

The renamed Northeastern Illinois Regional Crime Laboratory (NIRCL) used sex chromosome DNA profile (Y-STR analysis) on several items of evidence to try and obtain additional genetic identity information. This work also excluded Mr. Williford as a contributor.

Later forensic analysis

Several later laboratory reports provide an additional perspective on this case.

An NIRCL report, issued in this intermediate time frame, describes the results of a DNA database search conducted with a DNA profile derived from a re-extraction (*i.e.*, a second laboratory processing on the same item of evidence) from item 01.01, wood.

This DNA profile identified a probative sample from an unsolved Lake County Illinois cases and a DNA "hit" to the DNA found in the "Pooled Spermatozoa from Holly Staker Vaginal Swabs."

The DNA testing in this intermediate period also identified a partial profile "Stain D" on the same item of evidence 01.01, wood. Although the data from stain D is sparse, this same DNA profile, (i.e., this same contributor) appears to be present on other tested areas of item 01.01, wood.

Another item of evidence in this case, item 2 (red gas can) thought to have been the source of the accelerant used in the fire at the crime scene which may have been set to possibly try and cover up the homicide.

The analysis of this item revealed a relatively robust male DNA profile from the handle of the gas can.

Mr. Williford is excluded as a contributor to the DNA profile recovered from this item.

Recent forensic analysis

Here the previously obtained DNA profiles from (all generated previously by NIRCL) were analyzed using a recent DNA profile analysis software package / program, STRmix. More technically this program is a continuous probabilistic mixture interpretation software package.

This software program was developed to try and provide analysis of mixed DNA profiles; *i.e.*, DNA profiles produced when there are multiple contributors on an item of evidence.

Mixed DNA profiles are a regular feature in forensic DNA testing and arguably the deconvolution and analysis of mixed DNA profiles can be both complicated and controversial.

The STRmix program attempts to address this issue.

The software develops a series of iterative computer models that eventually approximate the DNA profile data that the laboratory records from the observed mixed /multiple contributor profile.

The model calculation is an iterative computation process that (for successful examples) arrives at what the software considers its 'best' model. As this is a computer generated DNA profile, the probability of each of the possible genotypes at the tested loci can be calculated.

The software then compares this model to a reference profile, typically the profile of a defendant.

Given the computed model and a reference profile, the software calculates a statistic, a likelihood ratio, for the relative strength of two (2) mutually exclusive hypotheses.

In general, one hypothesis would state that the mixed DNA profile is best modeled by the presence of the reference profile and additional unknown contributors.

Put another way, this hypothesis would claim that the defendant's profile is possibly present.

In general the opposing hypothesis would claim that the mixed DNA profile is best modeled by the presence of only unknown contributors, *i.e.*, that the reference profile is not present in the mixed DNA profile.

Put another way, this hypothesis would claim that the defendant is excluded.

The mathematical ratio between the probability of these two hypothesizes is a likelihood ratio: the relative strength of the of these two mutually exclusive outcomes:

(a) the defendant and some unknown individual(s) are present in the mixture

versus

(b) only unknown individual(s) are present in the mixed DNA profile.

The likelihood ratio is a pure number – there are no units in a likelihood ratio – it is just the odds of the probability of one hypothesis compared to the probability of another (mutually exclusive), hypothesis.

Due to the math of a likelihood ratio, there are three (3) overall possible results:

a) the likelihood ratio is larger than one (>1),

b) the likelihood ratio is around one (~1),

or

c) the likelihood ratio is smaller than one (<1).

A likelihood ratio larger than one would favor one hypothesis over its opposite.

A likelihood ratio smaller than one would favor the opposite hypothesis.

A likelihood ratio of around one (~1) demonstrates that the analysis does not provide any insight into the relative probabilities of the two hypotheses.

Other types of analysis (and other software packages) can certainly provide different results and there are known examples where this occurs.

What is particularly important is the context for the likelihood ratio: how much larger than one (>1) is meaningful for attempting to claim identity.

Interestingly this issue (a contentious one in forensic DNA), is not relevant here as the STRmix analysis excludes Mr. Williford from all of the mixed DNA profiles that were analyzed by STRmix.

The decedent victim, previously unambiguously identified, is also recognized by the STRmix analysis.

More specifically, the STRmix software analysis conducted by NIRCL excludes Mr. Williford as a contributor to item 26.01A (duct tape stain A).

More specifically, the STRmix software analysis conducted by NIRCL excludes Mr. Williford as a contributor to item 15.01D (right boot stain D).

More specifically, the STRmix software analysis conducted by NIRCL excludes Mr. Williford as a contributor to item 01.01B (wood stain B).

More specifically, the STRmix software analysis conducted by NIRCL excludes Mr. Williford as a contributor to item 01.01D (wood stain D).

More specifically, the STRmix software analysis conducted by NIRCL excludes Mr. Williford as a contributor to item 02.01C (gas can stain C

More specifically, the STRmix software analysis conducted by NIRCL excludes Mr. Williford as a contributor to item 02.01E (gas can stain E).

More specifically, the STRmix software analysis conducted by NIRCL excludes Mr. Williford as a contributor to item 01.01BE2 (wood stain B, second extraction).

More specifically, the STRmix software analysis conducted by NIRCL excludes Mr. Williford as a contributor to item 01.01GE2 (wood stain G, second extraction).

Put another way, a repetition of the exclusion of Mr. Williford as a contributor to tested samples in this case eight (8) more times.

It can be noted that NIRCL excluded the defendant from some of these samples previously and that this (correct) exclusionary conclusion was obtained without any software analysis whatsoever.

The computer software package STRmix was designed specifically to analyze DNA mixtures with the goal of removing (or at least minimizing) the effect of pre-formed bias.

In other words, the most recent DNA testing and analysis using STRmix, provides the most recent and crime laboratory acceptable analysis of complex mixtures.

This analysis was not available at the time of trial and could not have been presented to the jury as neither the updated DNA testing kits nor the STRmix software existed at the time. The STRmix analysis was also not available at the time of the ca. 2014 DNA testing.

It is important not to be confused by the 'verbal equivalent' that the NIRCl report uses in its statements.

The verbal equivalent is not a part of the likelihood ratio, is not provided by the STRmix software, is not a description of the number and is both arbitrary, unnecessary and has no scientific foundation.

The likelihood ratio does not contain a verbal description, is not described by words and can only be compared to another number. The verbal equivalent is to be ignored as it cannot and does not provide any insight into the likelihood ratio.

Here, all of the calculated likelihood ratios for all of the listed samples are smaller than one ((<1) and thus the software has calculated that the opposite hypothesis (*i.e.*, Mr. Williford is not one of the contributors to the mixed DNA profile) is more probable.

Conclusion

As was forecast previously in this document, Mr. Williford is excluded as a contributor from all evidence items that have been analyzed in this case.

This includes the wood (the club-like object used in the assault), the gas can, another probative item evidence used in the assault and all other items that produced DNA profiles that were analyzed.

Not only is Mr. Williford's DNA nowhere to be found (on anything), clear, unambiguous DNA evidence is present that identifies other individuals whose DNA was found on the items of evidence, often mixed with the victim's DNA (item 01.01, stain D, Unidentified Male #1, and from the item 02, gas can)

Karl Reich, Ph.D.
Chief Scientific Officer / Laboratory Director
Managing Partner
Independent Forensics
Lombard IL

EXHIBIT 8

Northeastern Illinois Regional Crime Laboratory

CODIS Report

1000 Butterfield Road, Suite 1009, Vernon Hills, IL 60061
Phone: (847) 362-0676 Fax: (847) 362-0712

Board President Robert LaMantia
Executive Director Garth Glassburg

Chief James Jackson
North Chicago Police Department
1850 Lewis Avenue
North Chicago, IL 60064

Subject: Homicide Investigation
Agency Case #: 00-002716
Case Officer: SAL CECALA
Submission Date: 01/17/2014

Laboratory Case #: 00-483
Laboratory Report #: 14
Report Date: 06/05/2014

Case Names: DELWIN NMI FOXWORTH SR
MARVIN T WILLFORD

The following evidence was submitted/retained in a sealed condition:

ITEM 01 (1SC) Exhibit 01 **extracted DNA from stains from wood (second extraction, major profile)**

RESULTS

A consistent DNA profile was identified in a search of the Combined DNA Index System (CODIS). The search detected a possible match between the DNA profile obtained from the above retained stain (report 13) and Lake County Sheriff's Office, Case #92-55313, Item #SMK002 (sample 2, major profile).

For informational purposes, this profile is also a possible match with the DNA profile obtained from "Pooled Spermatozoa from Holly Staker Vaginal Swabs" developed at Forensic Science Associates (File #05-001, refer to report 6).

If you have any questions, please contact this examiner.

The results portion of this report contains scientific judgments and interpretations rendered by the forensic scientist whose signature appears on the report.

Sarah E. Owen
Reviewer

M Kelly G Lawrence
Forensic Scientist
Kelly G. Lawrence

Northeastern Illinois Regional Crime Laboratory

DNA Report

1000 Butterfield Road, Suite 1009, Vernon Hills, IL 60061
Phone: (847) 362-0676 Fax: (847) 362-0712

Board President
Robert LaMantia

Executive Director
Garth Glassburg

Chief James Jackson
North Chicago Police Department
1850 Lewis Avenue
North Chicago, IL 60064

Subject: Homicide Investigation
Agency Case #: 00-002716
Case Officer: SAL CECALA
Submission Date: 12/10/2013, 03/23/2004, 01/17/2014, 04/08/2004

Laboratory Case #: 00-483
Laboratory Report #: 13
Report Date: 05/13/2014

Case Names: DELWIN NMI FOXWORTH SR
MARVIN T WILLFORD

The following evidence was submitted/retained in a sealed condition:

ITEM 01 (1SC)	Exhibit 01	retained stains from wood (report 11)
ITEM 15 (7SC)	Exhibit 01	pair of boots
ITEM 21	Exhibit 01	duct tape
ITEM 22	Exhibit 01	two pieces of duct tape
ITEM 26 (4DB)	Exhibit 01	wad of duct tape
ITEM 29 (7DB)	Exhibit 01	retained stain A from sweatshirt (report 10)
ITEM 36 (12DB)	Exhibit 01	retained standard of "John Adams"
ITEM 37 (13DB)	Exhibit 01	retained standard from "Amarin Willford"
ITEM 38 (11DB)	Exhibit 01	retained standard from "Delwin L. Foxworth"

RESULTS

Blood was indicated on duct tape (26.01).

Stains collected from right boot (15.01, stains A, C, and D), left boot (15.01, stains B, E, and F), duct tape (21.01, stains A and B), duct tape (22.01, stains A and B), duct tape (26.01, stain A) were extracted for DNA.

The above retained stains and standards were re-extracted for DNA.

The stains from wood (01.01, second extraction), right boot (15.01, stains A, C, and D), left boot (15.01, stains B, E, and F), duct tape (21.01, stains A and B), duct tape (22.01, stains A

Lab Case 00-483
Lab Report # 13
Analyst Kelly G. Lawrence

and B) and duct tape (26.01, stain A) are of human origin. DNA profiling was conducted by PCR using the Identifiler Plus amplification kit 15 STR loci and Amelogenin.

Partial DNA profiles originating from at least one unknown contributor were obtained from right boot (15.01, stain A) and duct tape (21.01, stains A and B). These profiles were not suitable for comparison analysis.

Partial DNA profiles originating from at least two unknown contributors were obtained from duct tape (22.01, stains A and B) and duct tape (26.01, stain A). These profiles were not suitable for comparison analysis.

A DNA profile was obtained from left boot (15.01, stain B) that matches the DNA profile obtained from "Delwin L. Foxworth" (second extraction).

"John Adams" (second extraction) and "Amarin Willford" (second extraction) are excluded as contributors of the DNA.

Partial DNA profiles were obtained from right boot (15.01, stain C) and left boot (15.01, stains E and F) that are consistent with coming from "Delwin L. Foxworth" (second extraction).

"John Adams" (second extraction) and "Amarin Willford" (second extraction) are excluded as contributors of the DNA.

A partial DNA profile originating from at least three unknown contributors was obtained from right boot (15.01, stain D). This profile was not suitable for comparison analysis.

A DNA profile originating from at least two contributors was obtained from wood (01.01, second extraction).

Assuming only two contributors are present, and "Delwin L. Foxworth" (second extraction) is the contributor of the minor profile, a major profile was deduced. The major profile is consistent with coming from one unknown male contributor.

"John Adams" (second extraction) and "Amarin Willford" (second extraction) are excluded as contributors of the DNA.

The major profile will be entered into the Combined DNA Index System (CODIS) and will be periodically searched against other profiles in the system. Should any consistent DNA profiles be identified, your agency will be contacted.

Sweatshirt (29.01, stain A, second extraction) failed to yield a sufficient amount of DNA for comparison analysis.

DNA profiling of wood (01.01, second extraction), right boot (15.01, stains C and D), left boot (15.01, stains E and F) and the above standards was conducted by PCR using the Yfiler amplification kit 17 STR loci.

Partial Y haplotypes were obtained from right boot (15.01, stain C) and left boot (stains E and F) that are consistent with coming from "Delwin L. Foxworth" (second extraction).

"John Adams" (second extraction) and "Amarin Willford" (second extraction) are excluded as contributors of these Y haplotypes.

Lab Case 00-483
Lab Report # 13
Analyst Kelly G. Lawrence

A partial Y haplotype originating from more than one male was obtained from right boot (15.01, stain D). This Y haplotype is not suitable for comparison analysis.

A Y haplotype originating from more than one male was obtained from wood (01.01, second extraction). A major Y haplotype was deduced and is suitable for comparison analysis. A minor haplotype suitable for comparison could not be deduced.

"John Adams", "Amarin Willford" and "Delwin L. Foxworth" are excluded as contributors to the major Y haplotype.

DNA evidence will be maintained at the laboratory should further analysis be required.

Analysis was not conducted on item 02 (16SC).

Should further analysis be required, please contact this examiner.

The results portion of this report contains scientific judgments and interpretations rendered by the forensic scientist whose signature appears on the report.

Please pick up all appropriate exhibits at your earliest convenience.

Sarah E. Owen
Reviewer

Kelly G. Lawrence
Forensic Scientist
Kelly G. Lawrence

Northeastern Illinois Regional Crime Laboratory

DNA Report

1000 Butterfield Road, Suite 1009, Vernon Hills, IL 60061
Phone: (847) 362-0676 Fax: (847) 362-0712

Board President Robert LaMantia
Executive Director Garth Glassburg

Chief James Jackson
North Chicago Police Department
1850 Lewis Avenue
North Chicago, IL 60064

Subject: Homicide Investigation
Agency Case #: 00-002716
Case Officer: SAL CECALA
Submission Date: 07/23/2014, 08/15/2003

Laboratory Case #: 00-483
Laboratory Report #: 17
Report Date: 12/02/2014

Case Names: DELWIN NMI FOXWORTH SR
MARVIN T WILLFORD

The following evidence was submitted/retained in a sealed condition:

ITEM 04 (18CC)	Exhibit 01	one cordless phone
ITEM 11 (3SC)	Exhibit 01	one baby wipes container (previously marked 11.09)
	Exhibit 02	one broken cd (previously marked 11.08)
	Exhibit 03	one "Crown Royal" glass (previously marked 11.02)
	Exhibit 04	one "Crown Royal" glass (previously marked 11.01)
	Exhibit 05	one wooden box (previously marked 11.05)
	Exhibit 06	one liquid soap dispenser (previously marked 11.06)
	Exhibit 07	one package of diapers (previously marked 11.07)
	Exhibit 08	one crystal bowl (previously marked 11.03)
ITEM 14 (6SC)	Exhibit 01	one black wallet
	Exhibit 02	one sealed envelope containing business cards and a plastic photo album
ITEM 17 (9SC)	Exhibit 01	one table lamp in three pieces
ITEM 19 (15CS)	Exhibit 01	one matchbook
ITEM 25 (3DB)	Exhibit 01	one purple lighter
ITEM 39 (6SC A)	Exhibit 01	one "wallet chain"
	Exhibit 02	one pair of clip-on earrings, one of which has a plastic piece by clip

Lab Case 00-483
Lab Report # 17
Analyst Kelly G. Lawrence

RESULTS

Blood was indicated on baby wipes container (11.01, stain A), soap dispenser (11.06), diapers (11.07), and lamp (17.01, stain C).

A visual analysis failed to detect any blood-like stains on wooden box (11.05).

The stains collected from cordless phone (04.01, stains A, B and C), baby wipes container (11.01, stain A), cd (11.02), glass (11.03), glass (11.04), wooden box (11.05), soap dispenser (11.06, stains A and B), crystal bowl (11.08), inside wallet (14.01, stain A) outside wallet (14.01, stain B), lamp (17.01, stains A, B, C and D), matchbook (19.01), lighter (25.01), wallet chain (39.01), earrings (39.02, stain A) and plastic piece (39.02, stain B) were extracted for DNA.

The stains from cordless phone (04.01, stains A, B, and C), cd (11.02), glass (11.03), glass (11.04), wooden box (11.05), soap dispenser (11.06, stains A and B), crystal bowl (11.08), inside wallet (14.01, stain A) outside wallet (14.01, stain B), lamp (17.01, stains A, B, and C), matchbook (19.01), lighter (25.01), wallet chain (39.01) and earrings (39.02, stain A) and plastic piece (39.02, stain B) are of human origin.

DNA profiling was conducted by PCR using the Identifiler Plus amplification kit 15 STR loci and Amelogenin.

A partial DNA profile was obtained from glass (11.03) that is consistent with coming from "Delwin L. Foxworth" (second extraction).

"John Adams" (second extraction), "Amarin Willford" (second extraction), and the contributor of the major profile obtained from wood (01.01, second extraction, report 13) are excluded as contributors of the DNA.

A partial DNA profile originating from at least two unknown contributors was obtained from inside wallet (14.01, stain A). This profile was not suitable for comparison analysis.

Cordless phone (04.01, stains A, B and C), baby wipes container (11.01, stain A), cd (11.02), glass (11.04), wooden box (11.05), soap dispenser (11.06, stains A and B), crystal bowl (11.08), stain A) outside wallet (14.01, stain B), lamp (17.01, stains A, B, C and D), matchbook (19.01), lighter (25.01), wallet chain (39.01) and earrings (39.02, stain A) and plastic piece (39.02, stain B) failed to yield a sufficient amount of DNA for comparison analysis.

DNA evidence will be maintained at the laboratory should further analysis be required.

Additional analysis was not conducted on diapers (11.07).

Analysis was not conducted on business cards and photo album (14.02).

For informational purposes, the DNA profile obtained from wood (01.01, first extraction, report 11) was re-interpreted with current interpretation guidelines. This profile is no longer suitable for comparison analysis.

Lab Case 00-483
Lab Report # 17
Analyst Kelly G. Lawrence

Should further analysis be required, please contact this examiner.

The results portion of this report contains scientific judgments and interpretations rendered by the forensic scientist whose signature appears on the report.

Please pick up all appropriate exhibits at your earliest convenience.

Reviewer

Forensic Scientist
Kelly G. Lawrence

Northeastern Illinois Regional Crime Laboratory

DNA Report

1000 Butterfield Road, Suite 1009, Vernon Hills, IL 60061
Phone: (847) 362-0676 Fax: (847) 362-0712

Board President
Robert LaMantia

Executive Director
Garth Glassburg

aChief James Jackson
North Chicago Police Department
1850 Lewis Avenue
North Chicago, IL 60064

Subject: Homicide Investigation
Agency Case #: 00-002716
Case Officer: SAL CECALA
Submission Date: 07/23/2014, 01/17/2014

Laboratory Case #: 00-483
Laboratory Report #: 18
Report Date: 01/07/2015

Case Names: DELWIN NMI FOXWORTH SR
MARVIN T WILLFORD

The following evidence was submitted/retained in a sealed condition:

ITEM 01 (1SC) Exhibit 01 piece of wood:

1. Extracted DNA from swab 1, first extraction (report 11)
2. Extracted DNA from swab 1, second extraction (report 13)
3. Retained swab 1 (report 11)
4. Collected stains, swab 2
5. Collected stains, swab 3
6. Collected stains, swab 4
7. Retained stain A, F (report 5)
8. Retained stain B (report 5)
9. Retained stain D (report 5)
10. Retained stain E, I (report 5)
11. Retained stain G (report 5)
12. Retained stain H (report 5)
13. Retained stain J (report 5)
14. Retained stain K (report 5)

ITEM 29 (7DB) Exhibit 01 retained stain A from sweatshirt (report 10)

RESULTS

Human blood was detected on sweatshirt (29.01, stain A).

The stains from wood (swab 2, 3, 4 and stains A,F, B, D, E,I, G, H J, and K) were extracted for DNA.

Stains from wood (01.01, swab 1) and sweatshirt (29.01, stain A) were re-extracted (third extraction) for DNA.

These stains are of human origin.

Lab Case 00-483
Lab Report # 18
Analyst Kelly G. Lawrence

The extracted DNA from wood (01.01, swab 1, first and second extractions) were re-amplified.

DNA profiling of the above stains was conducted by PCR using the Identifiler Plus amplification kit 15 STR loci and Amelogenin.

DNA profiles each originating from at least three unknown contributors were obtained from wood (01.01, swab 1, third extraction), wood (01.01, swab 3), and wood (01.01, swab 4). These profiles were not suitable for comparison analysis.

A DNA profile was obtained from wood (01.01, stains A,F, E,I, H, J, and K) that matches the DNA profile obtained from "Delwin L. Foxworth". "John Adams" and "Amarin Willford" are excluded as contributors of the DNA.

A DNA profile originating from at least two contributors was obtained from wood (01.01, stain B). Assuming only two contributors are present and "Delwin L. Foxworth" is a contributor to the mixture, a second profile suitable for comparison analysis could not be deduced.

A partial DNA profile originating from at least two contributors was obtained from wood (01.01, stain D). Assuming only two contributors are present and "Delwin L. Foxworth" is a contributor to the mixture, a second profile was deduced. The deduced profile is consistent with coming from one unknown contributor.

This deduced profile is suitable for comparison analysis with known standards, but is not suitable for entry into the Combined DNA Index System (CODIS).

"John Adams" and "Amarin Willford" are excluded as contributors of the deduced profile.

A DNA profile originating from at least two contributors was obtained from wood (01.01, stain G). Assuming only two contributors are present and "Delwin L. Foxworth" is a contributor to the mixture, a second profile suitable for comparison analysis could not be deduced.

A partial DNA profile originating from at least two contributors was obtained from wood (01.01, swab 1, first extraction). Assuming only two contributors are present and "Delwin L. Foxworth" is a contributor to the mixture, a second profile suitable for comparison analysis could not be deduced.

A DNA profile originating from at least two contributors was obtained from wood (01.01, swab 1, second extraction). Assuming only two contributors are present and "Delwin L. Foxworth" is a minor contributor to the mixture, a major profile was deduced. The major profile is consistent with coming from one unknown male contributor.

The contributor of the major profile obtained from wood (01.01, swab 1, second extraction, report 13) cannot be excluded as the contributor of the major profile.

"John Adams", "Amarin Willford", and the contributor of the DNA profile obtained from gas can handle (Independent Forensics Case #27730-80967, sample Q1) are excluded as contributors of the major profile.

The stains from wood (01.01, swab 2) and sweatshirt (29.01, stain A, third extraction) failed to yield a sufficient amount of DNA for comparison analysis.

Lab Case 00-483
Lab Report # 18
Analyst Kelly G. Lawrence

DNA evidence will be maintained at the laboratory should further analysis be required.

Additional analysis of wood (01.01, stains B and G) and matchbook (19.01) will be conducted. The results of that analysis will be the subject of a subsequent report.

Should further analysis be required, please contact this examiner.

The results portion of this report contains scientific judgments and interpretations rendered by the forensic scientist whose signature appears on the report.

Please pick up all appropriate exhibits at your earliest convenience.

Sarah E. Owen
Reviewer

Kelly G. Lawrence
Forensic Scientist
Kelly G. Lawrence

Northeastern Illinois Regional Crime Laboratory

DNA Report

1000 Butterfield Road, Suite 1009, Vernon Hills, IL 60061
Phone: (847) 362-0676 Fax: (847) 362-0712

Board President
Robert LaMantia

Executive Director
Garth Glassburg

Chief James Jackson
North Chicago Police Department
1850 Lewis Avenue
North Chicago, IL 60064

Subject: Homicide Investigation
Agency Case #: 00-002716
Case Officer: SAL CECALA
Submission Date: 07/23/2014, 01/17/2014

Laboratory Case #: 00-483
Laboratory Report #: 19
Report Date: 01/09/2015

Case Names: DELWIN NMI FOXWORTH SR
MARVIN T WILLFORD

The following evidence was submitted/retained in a sealed condition:

ITEM 01 (1SC)	Exhibit 01 wood, stains B and G (report 5)
ITEM 02 (16SC)	Exhibit 01 one gas can
ITEM 18 (10SC)	Exhibit 01 one pair of black shoes
	Exhibit 02 one pair of "Air Jordan" shoes
ITEM 19 (15CS)	Exhibit 01 one matchbook, of which various cuttings were tested (stain B)
ITEM 20 (LT1)	Exhibit 01 one glass vial, pair of pants, and one pair of underwear, of which the ankle areas of each leg of pants (stain A) was tested
ITEM 27 (5DB)	Exhibit 01 one shirt, of which the wrist areas of each arm cuff (stains A and B) were tested
ITEM 28 (6DB)	Exhibit 01 one pair of sweatpants, of which the ankle areas of each leg (stains B and C) were tested
ITEM 29 (7DB)	Exhibit 01 one sweatshirt, of which the outside wrist cuffs of each arm (stains C and D) were tested

RESULTS

Blood was indicated on "Air Jordan" shoes (18.02).

Analysis for blood failed to yield conclusive results on stains from pants (20.01).

Lab Case 00-483
Lab Report # 19
Analyst Kelly G. Lawrence

The stains collected from gas can (02.01, stains A, B, C, D, E and F), black shoes (18.01, stains A and B), "Air Jordan" shoes (18.02, stains A, B, C and D), matchbook (19.01, stain B), pants (20.01, stain A), shirt (27.01, stains A and B), sweatpants (28.01, stains B and C) and sweatshirt (29.01, stains C and D) were extracted for DNA.

The stains from wood (01.01, stains B and G) were re-extracted (second extraction for each) for DNA.

These stains are of human origin. DNA profiling was conducted by PCR using the Identifiler Plus amplification kit 15 STR loci and Amelogenin.

Partial DNA profiles each originating from at least two unknown contributors were obtained from gas can (02.01, stains A, C, E) and pants (20.01, stain A). These profiles were not suitable for comparison analysis.

A partial DNA profile originating from at least one unknown contributor was obtained from gas can (02.01, stain F). This profile was not suitable for comparison analysis.

A DNA profile originating from at least two contributors was obtained from wood (01.01, stain B, second extraction). Assuming only two contributors are present, a major profile and minor profile were deduced.

The major profile is consistent with coming from "Delwin L. Foxworth".

The minor profile is consistent with coming from one unknown contributor. This profile is suitable for comparison analysis with known standards, but is not suitable for entry into the Combined DNA Index System (CODIS).

"John Adams", "Amarin Willford" and the contributors of the DNA profiles obtained from wood (01.01, second extraction, major profile, first and second amplifications, reports 13 and 18), wood (01.01, stain D, deduced profile, report 18) and gas can handle (Independent Forensics Case #27730-80967, sample Q1) can be excluded as contributors of the major profile.

"John Adams", "Amarin Willford", and "Delwin L. Foxworth" can be excluded as contributors of the minor profile.

A DNA profile originating from at least two contributors was obtained from wood (01.01, stain G, second extraction). Assuming only two contributors are present, a major profile and minor profile were deduced.

The major profile is consistent with coming from "Delwin L. Foxworth".

The minor profile is consistent with coming from one unknown contributor. This profile is suitable for comparison analysis with known standards, but is not suitable for entry into the Combined DNA Index System (CODIS).

"John Adams", "Amarin Willford" and the contributors of the DNA profiles obtained from wood (01.01, second extraction, major profile, first and second amplifications, reports 13 and 18), wood (01.01, stain D, deduced profile, report 18) and gas can handle (Independent Forensics Case #27730-80967, sample Q1) can be excluded as contributors of the major profile.

Lab Case 00-483
Lab Report # 19
Analyst Kelly G. Lawrence

"John Adams", "Amarin Willford", and "Delwin L. Foxworth" can be excluded as contributors of the minor profile.

The stains from gas can (02.01, stains B and D), shoes (18.01, stains A and B), shoes (18.02, stains A, B, C and D), matchbook (19.01, stain B), shirt (27.01, stains A and B), sweatpants (28.01, stains B and C) and sweatshirt (29.01, stain C and D) failed to yield a sufficient amount of DNA for comparison analysis.

DNA evidence will be maintained at the laboratory should further analysis be required.

The stain(s) collected from pants (20.01, stain A) were retained at the laboratory.

Should further analysis be required, please contact this examiner.

The results portion of this report contains scientific judgments and interpretations rendered by the forensic scientist whose signature appears on the report.

Please pick up all appropriate exhibits at your earliest convenience.

Sarah E. Owen
Reviewer

M Kelly G Lawrence
Forensic Scientist
Kelly G. Lawrence

EXHIBIT 9

IN THE CIRCUIT COURT OF LAKE COUNTY
ILLINOIS COUNTY DEPARTMENT, LAW DIVISION

State of Illinois,	)	
	)	
Respondent,	)	
	)	
vs.	)	No. 00 CF 1920
	)	
MARVIN T. WILLIFORD	)	
Petitioner	)	
	)	
	)	
	)	

AFFIDAVIT OF DR. KARL REICH

I, Dr. Karl Reich, under oath and penalty of perjury, subscribe and swear as follows:

1. I have a doctorate in Molecular Biology and am the Chief Scientific Officer of Independent Forensics of Illinois, 500 Waters Edge, Suite 210, Lombard, IL 60148. A true and correct copy of my *curriculum vitae* is attached as Ex. A to this affidavit.

2. Independent Forensics Laboratory adheres to the FBI's Quality Assurance Standards for molecular biology, human genetics and forensic DNA Testing Laboratories. Independent Forensics is accredited by Forensic Quality Services-International (FQS-I, ISO/IEC 17025), the American Association of Blood Banks (AABB), and the New York State Department of Health (NY-DOH), ~~for genetic identity and forensic~~ DNA testing.

3. Independent Forensics subscribes to the College of American Pathologists (CAP) and to Collaborative Testing Services (CTS) for external proficiency testing programs for Quality Assurance and Quality Control for Forensic DNA analysis and Disputed Parentage/ Family Relationship Testing.

4. I am very familiar with the scientific literature, research efforts and technologies used to analyze and interpret DNA-based forensic evidence. This familiarity derives from the hands-on analysis of hundreds of DNA samples in a fully accredited DNA forensic laboratory, from supervising the development of new forensic tests that are used by forensic laboratories

world-wide and from the professional review of several hundred DNA-based cases.

5. I have been asked by counsel for the Petitioner to assess, critically evaluate, review and summarize the forensic DNA results obtained to date in this case. The list of documents and materials reviewed in this process is attached as Ex. B to this affidavit.

6. The Northeastern Illinois Regional Crime Laboratory Case (NIRCL) used generally accepted methods and procedures to process the many items of evidence tested in this case. It should be noted that the range of evidence tested includes (potentially) touched or handled evidence, apparent stains, and of course reference or known samples from the decedent and defendant.

7. There are several examples where NIRCL apparently re-extracted evidence samples. There is no prohibition against going back to previously examined and processed evidence to try and extract more biological material; however there is only a finite amount of biological material originally deposited on the item and re-extraction does not manufacture additional DNA.

8. Forensic DNA laboratories are governed by their standard operating procedures (SOPs) where not only the day to day work of the laboratory is described and codified, but also the interpretation of the collected data is delineated. The laboratory is bound by its SOP strictures - external case review is not. External review of data can reveal different conclusions, summarize more, or less data than the laboratory and can be scientifically reliable, accepted and appropriate. External case review can take advantage of a more nuanced and experienced perspective and provides a true peer review of the forensic DNA evidence in a case.

9. Item of evidence:
   Wood, stains on wood, extractions from wood, second extractions from wood. This item of evidence was apparently used to physically strike the decedent and has been extensively examined and processed for forensic DNA. At least fourteen different amplifications were performed on various stains, sections and extractions derived from this item of evidence.

   Mr. Williford is excluded as a contributor from any and all DNA profiles (deemed suitable for comparison by NIRCL) derived from item 01.01, wood. It is important to note that in the DNA profiles that were not deemed suitable for comparison, there are no data or even a hint of data linked to Mr. Williford.

In most, but not all, of the DNA profiles obtained from this item, the DNA profile from the decedent, Delwin Foxworth can be observed.
Many of the observed DNA profiles from this item of evidence are mixed DNA profiles, derived from at least two, contributors. None of the possible contributors from these mixed DNA profiles is Mr. Williford.

A mixed DNA profile derived from the wood, 01.01 swb E2 A2 was appropriately analyzed to deduce a major and a minor profile. Here major refers exclusively to the subset of peaks from the electropherogram that are larger (greater signal intensity). Here minor refers exclusively to the subset of peaks from the electropherogram that are smaller (less signal intensity). No inference to importance, order or timing of DNA contribution can be ascribed to the terms major or minor – these refer exclusively to the relative amounts of DNA from the two contributors. The major DNA profile from this sample was used in a DNA database search and is unambiguously linked to another, unsolved, Lake county crime, specifically the 'Holly Staker Case', 'pooled spermatozoa from Holly Staker vaginal swabs' referred here to as 'Unidentified Male#1'. The minor contributor to this sample was the decedent.

A deduced partial DNA profile was obtained from stain D identified on this exhibit. Results from seven (7) loci were reported – Mr. Williford is excluded as a contributor to this partial profile, the decedent is excluded, the new DNA profile from the gas can is excluded and the DNA profile listed as 'Unidentified Male#1' is also excluded as a contributor.

A mixed DNA profile from wood, 01.01 BE2, was appropriately analyzed to deduce a major and minor profile. The major profile is derived from the decedent, while the minor profile is consistent with the partial profile deduced from stain D; Mr. Williford is excluded as a contributor to this stain. The new DNA profile from the gas can is excluded and the DNA profile listed as 'Unidentified Male#1' is also excluded as a contributor.

A mixed DNA profile from wood, 01.01 GE2, has at least two contributors, one of whom must be male. One possible contributor is consistent with the decedent and the other possible contributor is consistent with the partial profile deduced from stain D; Mr. Williford is excluded as a contributor to this stain. The new DNA profile from the gas can is excluded and the DNA profile listed as 'Unidentified Male#1' is also excluded as a contributor.

Technically the data described above is derived from the following:
Item 01.01 A B, B E2, D, E, E2A2, E3, F, G, G E2, H, I, J, K. Wood SWB A2, SWB2 E2, SWB3, SWB4.

10. Item of evidence
    Duct tape, extractions from duct tape. This item of evidence was apparently used to bind the decedent. Only partial, incomplete DNA profiles were obtained from this item of evidence, some of these partial profiles were mixtures from at least two contributors. These minimal data can be accounted for by the decedent or someone else besides Mr. Williford.

    Mr. Williford is excluded as a contributor from all samples derived from duct tape. The new DNA profile from the gas can is excluded from 21.01A but cannot be excluded from item 21.01B. The DNA profile listed as 'Unidentified Male#1' is excluded as a contributor.

    The deduced DNA profile from stain D, exhibit 01.01 wood, cannot be excluded as a contributor to these exhibits.

    Technically the data described above is derived from the following:
    Item 21.01 A, B; Item 22.01 A, B; Item 26.01 A.

11. Item of evidence
    Gas can. This item of evidence was apparently used as the source of accelerant in the assault on Mr. Foxworth. Two laboratories, Independent Forensics, and NIRCL, examined and processed this item of evidence for DNA profiles.

    DNA profiles suitable for comparison were obtained from this item of evidence, and Mr. Williford is excluded as a contributor to this item of evidence. Additional low level, partial DNA profile data were obtained from this item of evidence. It is important to note that in the DNA profiles that were not deemed suitable for comparison, there are no data or even a hint of data linked to Mr. Williford.

    Technically the data described above is derived from the following:
    Item 02.01 A, B, C, D, E, F; Independent Forensics 27730-80967-Q1

12. Item of evidence
    Boot, right boot and boot, left boot. These items were tested for forensic DNA and either very partial profiles, complete profiles or a partial mixed DNA profile was recorded. The decedent's DNA profile can account for the majority of the data seen from these items of evidence. Mr. Williford is excluded as a contributor in all instances recorded from these items.

    Technically the data described above is derived from the following:

Item 15.01 A, B, C, D, E, F

13. Item of evidence

Shoes, pants, shirt, sweatpants, sweatshirt, matchbook, soap dispenser, crystal bowl, baby wipes container, diapers, lamp, wooden box, cordless phone, CD, glass, wallet, lighter, wallet chain, earrings, plastic piece. These items were examined, processed and analyzed for DNA profiles.

Mr. Williford is excluded from the few items that provided a DNA profile suitable for comparison. Little, or in some cases, no DNA profile data was obtained from the analysis of these items of evidence or DNA profiles not deemed suitable for comparison or DNA profiles strongly linked to the decedent were observed. It is important to note that in the DNA profiles that were not deemed suitable for comparison, there are no data or even a hint of data linked to Mr. Williford.

Technically the data described above is derived from the following:
Item 18.01 A, B; Item 18.01 A, B, C, D. Item 20.01 A. Item 27.01 A, B. Item 28.01 B, C E2. Item 29.01 A E2, C E2, D E2. Item 19.01 B. Item 11.06 B. Item 11.08. Item 11.01. Item 11.07. Item 17.01. Item 11.05. Item 04.01. Item 11.02. Item 11.03, Item 11.04. Item 14.01. Item 25.01. Item 39.01. Item 39.02 A B.

14. Conclusions

An unusually large number of items of evidence and individual stains and areas of these items of evidence were processed for forensic DNA analysis.

In every and all cases Mr. Williford was excluded as a contributor from any DNA profile suitable for comparison. It is important to note that even from the DNA profiles deemed too incomplete for comparison or a DNA database search, no inference, data or hint of data links to Mr. Williford.

15. Mr. Williford is excluded from items of evidence directly linked to the assault on Mr. Foxworth including the wood, gas can and duct tape.

16. Probative DNA profiles were obtained from these items of evidence, including a DNA profile linked to another, unsolved Lake County homicide. Other DNA profiles suitable for DNA database search were identified and did not return a 'hit', *i.e.*, no similar DNA profiles were found in the DNA database.

17. No information was provided as to the extent of the DNA database search, *i.e.*, whether both local and national DNA databases were searched against the identified DNA profiles.

18. At least one DNA profile, from an unknown contributor, was identified on two items of evidence providing a tentative intra-case link.

    As noted above, a case-to-case link was identified using a DNA profile obtained from the wood.

19. Although not explicitly described, an apparent case of contamination by NIRCL personnel is present. Due to the nature of forensic DNA analysis, contamination is a fact of life; how this contamination occurred and the transparency of the laboratory in dealing with this issue is particularly important. NIRCL did not distinguish itself in this regard.

20. Mr. Williford cannot be linked to any item of evidence in this case. He is excluded from all tested materials. No physical evidence links him to this case.

21. Probative DNA profiles were obtained by NIRCL, but the individuals who contributed these DNA profiles remain unknown.

Further this affiant sayeth naught.

Signed under the pains and penalties of perjury this 30th day of October, 2015.

Dr. Karl Reich, Ph.D.
Lombard, IL

STATE OF ILLINOIS
COUNTY OF DUPAGE

Sworn to and subscribed before me
this 30th day of October, 2015

Allan Suyosa Notary Public
My Commission Exp. March 30, 2018

ALLAN SUYOSA
OFFICIAL SEAL
Notary Public, State of Illinois
My Commission Expires
March 30, 2018

Exhibit A

Affidavit: Reich, Karl A, *curriculum vitae*

Karl A. Reich, Ph. D.

INDEPENDENT FORENSICS, 500 WATERS EDGE, SUITE 210 LOMBARD IL 60148
karl@IFI-Test.com; P 708.234.1200; F 708.978.5115

FORENSIC DNA / MOLECULAR BIOLOGY / MICROBIOLOGY / PROTEIN BIOCHEMISTRY MICROBIAL AND HUMAN FUNCTIONAL GENOMICS / PROTEIN PURIFICATION

Scientist with eight years post-graduate and fifteen years progressive experience in the pharmaceutical and biotechnology industry. Proven track record of initiating, managing and leading product oriented research in forensic DNA, genomics, infectious diseases, pharmaceutical target identification functional genomics, biotherapeutics, molecular biology, microbiology and strain development for industrial fermentation.

- Court Qualified DNA Expert Witness for Forensic DNA, Forensic Biology and Statistics – Testimony and depositions in more than fifty cases in State, Federal and International courts in both criminal and civil litigation.
- R & D project development and management from conception to market launch for forensic laboratory products.
- Developed, championed and implemented market-driven strategies for functional genomics biotech startup.
- R & D management experience in market-driven pharmaceutical, biotech and forensic DNA companies.
- R & D project development experience, including market analysis, target identification and validation, HTS, lead evaluation and animal efficacy trials.
- Led, built and managed research teams to implement strategic alliances, contract research and 'in-house' R & D in molecular biology, anti-infectives and strain development.

PROFESSIONAL EXPERIENCE

INDEPENDENT FORENSICS, Lombard IL **8/2002 – Present**

Chief Scientific Officer for DNA Forensics, Paternity, and Molecular Biology laboratory.

- Responsibility for development, validation, commercialization, production and manufacturing of new forensic-based tests.
- Supervisory responsibility for all laboratory operations, including validation, documentation, Q/A, Q/C, DNA testing, DNA analysis for forensics and paternity.
- Responsibility for lab design, lab set-up, IT, molecular biology, software and system design and implementation.
- PI on R & D contracts from federal law enforcement agency, PI on CDC SBIR grant, PI on DHS SBIR award.

ORCHID BIOSCIENCES, Long Island NY **4/2001 – 6/2002**

Pharmaceutical Development for 'virtual' pharmaceutical company.

- Responsibility for outsourcing of GMP synthesis of small molecule therapeutic compound.
- Initiated, negotiated and supervised CRO managed ongoing Phase II clinical trial.
- Supervised and outsourced FDA and EMEA fillings for Orphan Drug Status in Europe and U.S.A.
- Project fully acquired (and terminated) by strategic partner, 6/2002.

INTEGRATED GENOMICS, Chicago, IL **4/2001 – 2/2002**

Director of Pharmaceutical Development – Executive Management Team

Integrated Genomics is a startup functional genomics company focusing on a bioinformatic approach to solving industrial biotechnology problems.

- Responsibility for developing and implementing small molecule-based R & D for 'niche' anti-infectives markets.
- Developed research programs for strategic partners in anti-infective biology, industrial strain improvement, flavors and fragrance industries and genomic databases.
- Experience in, and responsible for, presenting research programs to pharmaceutical partners, venture capital funds and institutional investors.

DNA & IMMUNOGENETICS INSTITUTE, Chicago IL **6/2001 – 8/2002**

Co-Director, Laboratory Services

DNA & Immunogentics Institute was the first independent DNA testing laboratory in Illinois and performed testing for paternity determination, transplant matching and blood banks.

- Paternity & Forensic DNA Testing
- Blood Antigen Testing

ABBOTT LABORATORIES, Abbott Park IL **10/1996 – 4/2001**

Abbott Laboratories is a mid-tier pharmaceutical company with a strong focus on small molecule therapeutics.

- Directed, managed and led research group charged with cloning, expressing and purifying protein targets for pharmaceutical discovery and biotherapeutics using bacterial, insect and mammalian expression systems.
- Led effort to identify alternative expression systems/hosts for 'difficult' protein classes.
- Co-developed semi-automated cloning and expression system for HTS of protein targets.

Group Leader, Genomics and Molecular Biology.

- Devised, championed and directed all phases of genomics-based research program for the identification of novel anti-bacterial targets, including microbiology, mol bio, HTS cloning and expression, and database management.
- Identified and validated dozens of novel anti-bacterial targets.
- Conceived and managed small molecule discovery projects; including HTS assay development, hit characterization, animal efficacy models, SAR determination and toxicity profiles.
- Initiated proteomics program in *Haemophilus influenzae*.
- Developed and fabricated *H. influenzae* micro-array for inhibitor mode of action studies.
- Initiated and directed numerous external scientific collaborations.
- Developed broad knowledge base of genomic techniques, applications and technologies including SNPs, pharmacogenomics, proteomics, HTS sequencing, public and proprietary genomic databases.

STANFORD UNIVERSITY SCHOOL OF MEDICINE, Stanford CA **10/1990 - 10/1996**

Howard Hughes Research Fellow, Laboratory of Dr. Gary Schoolnik

- Developed research program on luminescent bacterial symbiont, *Vibrio. fischeri*.
- Discovered novel ADP-ribosyltransferase in culture supernatants of *V. fischeri*.
- Purified and cloned (using reverse genetics) novel ADP-ribosyltransferase from *V. fischeri*.
- Developed genetic system for *V. fischeri*, - targeted knock-outs for gene function identification.
- Initiated collaborative research with USC marine biology laboratory on symbiont/host interactions.

Post-doctoral fellow, Laboratory of Dr. Gary Schoolnik

- Developed research program on structure/function relationship of trans-membrane transcriptional activator, ToxR, in *V. cholerae*.
- Analyzed distribution of ToxR genes in environmental *Vibrio* isolates.
- Cloned, sequenced and characterized ToxR gene from luminescent marine bacterium, *V. fischeri*.

INSTITUT PASTEUR, Paris, France **10/1988 – 10/1990**

Fogarty Post-Doctoral Research Fellow

- Developed mono-clonal antibodies against membrane active toxin of *Listeria monocytogenes*.
- Developed novel, large scale purification protocol for listeriolysin.
- Participated in *in vivo* tests of single amino acid substituted *L. monocytogenes* isogenic strains.

PRIOR RELATED EMPLOYMENT

UCLA, Los Angeles CA
Dept of Biological Chemistry, Laboratory of Dr. D. Sigman **1979-1982**
Research Assistant
- Analysis of non-enzymatic cleavage of DNA by 1,10-orthophenanthroline Copper.
- Synthesized chemical derivatives of 1,10-orthophenanthroline.
- Recombinant over-expression and purification of *E. coli* DNA Polymerase.

HARVARD MEDICAL SCHOOL, Boston MA
Dept. of Neurobiology, Laboratory of Dr. T. Wiesel **1977-1979**
Research Assistant
- Developed micro-bore HPLC for amino acid analysis of retinal homogenates.
- General laboratory duties, including ordering, organization and solution preparation for histology and EM.

CORNELL UNIVERSITY, Ithaca, NY **summer 1976**
Laboratory of Dr. E. Ellson
Summer intern Low angle light scattering analysis of liposome preparations.

ROCKEFELLER UNIVERSITY, New York NY **summer 1975**
Laboratory of Dr. N. Zinder
Summer intern Production, purification and use of mini-cells as 'cell free' protein translation system.

EDUCATION

UCLA / HARVARD MEDICAL SCHOOL **1982-1988**
Ph.D. Molecular Biology
- ***Thesis: Enzymic Studies on Diphtheria Toxin Fragment A***

CORNELL UNIVERSITY **1973-1977**
B.A. Chemistry

LANGUAGES

- **ENGLISH, FRENCH**

PUBLICATIONS:

1. D.R. Graham, L.E. Marshall, **K.A. Reich** and D.S. Sigman, "Cleavage of DNA by Coordination Complexes. Superoxide Formation in the Oxidation of 1,10- Phenanthroline-Cuprous Complexes by Oxygen. Relevance to DNA-Cleavage Reaction," *J. Amer. Chem. Soc.*, **102**, 5419 (1980).

2. L.E. Marshall, D.R. Graham, **K.A. Reich** and D.S. Sigman, "Cleavage of DNA by the 1,10-Phenanthroline-Cuprous Complex. Hydrogen Peroxide Requirement and Primary and Secondary Structure Specificity," *Biochemistry*, **20**, 244 (1981).

3. **K.A. Reich**, L.E. Marshall, D.R. Graham and D.S. Sigman, "Cleavage of DNA by the1,10-Phenanthroline-Cuprous Complex. Superoxide Mediates the Reaction Dependent on NADH and Hydrogen Peroxide," *J. Amer. Chem. Soc.*, **103**, 3582 (1982).

4. L.M. Pope, **K.A. Reich**, D.R. Graham and D.S. Sigman, "Products of DNA Cleavage by the 1,10-Phenanthroline-Cuprous Complex. Product Analysis," *J. Biol. Chem.*, **257**, (20) 12121 (1982).

5. J.D.L. Harpe, **K.A. Reich**, E. Reich and E.B. Dowdle, "Diamphotoxin: The Arrow Poison of the !Kung Bushmen," *J. Biol. Chem.*, **258**, (19) 11924 (1983).

6. B.L. Kagan, **K.A. Reich** and R.J. Collier, "Orientation of the Diphtheria Toxin Channel in Lipid Bilayers," *Biophys. J.*, **45**, 102 (1984).

7. P. Berche, **K.A. Reich**, M. Bonnichon, J.L. Beretti, C. Geoffroy, J. Raveneau, P.Cossart, J.L. Gaillard, P. Geslin, H. Kreis and M. Veron, "Detection of Anti-Listeriolysin O for Serodiagnosis of Human Listeriosis," *Lancet*, **335**, 624-627 (1990).

8. B.A. Wilson, **K.A. Reich**, B.R. Weinstein and R.J. Collier, "Active-Site Mutations of Diphtheria Toxin: Effects of Replacing Glutamic Acid-148 with Aspartic Acid, Glutamine or Serine," *Biochemistry*, **29**, 8643-8651 (1990).

9. E. Michel, **K.A. Reich**, R. Favier, P. Berche and P. Cossart, "Attenuated Mutants of the Intracellular Bacterium *Listeria monocytogenes* obtained by Single Amino-Acid Substitutions in Listeriolysin O," *Molecular Microbiology*, **4** (12), 2167-2178 (1990).

10. F. Nato, **K.A. Reich**, S.Hopital, S. Rouyre, C. Geoffroy, J.C. Mazie and P. Cossart, "Production and Characterization of Monoclonal Antibodies against Listeriolysin O (LLO), the Thiol Activated Hemolysin of *Listeria monocytogenes*: Neutralizing Antibodies are Specific for LLO," *Infection and Immunity*, **59**(12), 4641-4646 (1991).

11. **K.A. Reich** and G. K. Schoolnik, "The Light Organ Symbiont *Vibrio fischeri* Possesses a Homologue of the *Vibrio cholerae* Transmembrane Transcriptional Activator ToxR", *J. Bacteriology*, **176**(10), 3085-3088 (1994).

12. B. Wilson, S.R. Blanke, **K.A. Reich** and R. John Collier, "Active-Site Mutations of Diphtheria Toxin. Tryptophan 50 is a Major Determinant of NAD Affinity. *J. Biol. Chem.*, **269**(37), 23296-23301, (1994)

13. **K.A. Reich** and G.K. Schoolnik, "Halovibrin, Secreted from the Light Organ Symbiont, *Vibrio fischeri*, Is a Member of a New Class of ADP-ribosyltransferases," *J. Bacteriology*, **178** (1), 209-215 (1996).

14. **K.A. Reich**, T. Biegel and G.K. Schoolnik, "The Light Organ Symbiont *Vibrio fischeri* Possesses Two Distinct Secreted ADP-ribosyltransferases," *J. Bacteriology*, **179**(5), 1591-1597 (1997).

15. E. P. Rock, **K. A. Reich**, D. M. Lyu, M. Hovi, J. Hardy, G. K. Schoonik, B. A. D. Stocker and V. Stevens. "Immunogenicity of LTB-hCG Fusion Protein," *Vaccine*, **14**(16), 1560-1568 (1996).

16. L. Katz, D. T. Chu and **K.A. Reich**. "Bacterial Genomics and the Search for Novel Antibiotics," *Annual Reports Med. Chem.*, **32**, 121-130, (1997).

17. P. Zhong, Z. Cao, R. Hammond, Y. Chen, J. Beyer V.D. Shortridge, L. Phan, S. Pratt, J. Capobianco, **K. A. Reich**, B. Flamm, Y. S. Or and L. Katz. "Induction Ribosome Methylation in MLS-resistant *Streptococcus pneumoniae* by macrolides and ketolides", *Microb Drug Resist*; **5**(3), 183-188, (1999).

18. **K. A. Reich**, P. Hessler, L. Chovan, "Genome Scanning in *Haemophilus influenzae*: *in vitro* transposition and PCR analysis for the identification of 'essential genes', *J. Bacteriology*, **181**(16), 4961-4968, (1999).

19. **K. A. Reich**, "The search for essential genes", *Res Microbiol* **151**(5), 319-324, (2000)

20. Stabb E. V., **K. A. Reich**, E. G. Ruby, "*Vibrio fischeri* genes hvnA and hvnB encode secreted NAD(+)-glycohydrolases", *J. Bacteriol.*, **183**(1), 309-317(2001).

21. Tack, L. C., M. Thomas, **K.A. Reich,** Automated Forensic DNA Purification Optimized for FTA Card Punches and Identifiler STR-based PCR Analysis, Clin Lab Medicine, 27(1), 183-9, 2007.

22. Johan H. Melendez, Julie A. Giles, Jeffrey D. Yuenger, Tukisa D. Smith, Khalil G. Ghanem, **Karl Reich** and Jonathan Zenilman. Detection and Quantification of Y Chromosomal Sequences by Real-Time PCR Using the LightCycler System, *Sexually Transmitted Diseases,* March 2007, Vol. 34, No. 3, 2007.

23. Brett Schweers, Jennifer Old, P.W. Boonlayangoor and **Karl Reich**, Developmental Validation of a Novel Lateral Flow Strip Test for Rapid Identification of Human Blood (Rapid Stain Identification™-Blood), Forensic Science International, 2, 243-247, 2008.

24. Jennifer Old, Brett Schweers, Jennifer Old, P.W. Boonlayangoor and **Karl Reich,** Developmental Validation of RSID™-Saliva; A Lateral Flow Immunochromatographic Strip Test for the Forensic Detection of Saliva, J Forensic Sci.,54 (4), 866-873, 2009.

25. Kevin W.P. Miller, Ph.D., Jennifer Old, Ph.D. , Brian R. Fischer, B.S., Brett Schweers, Ph.D. , Simona Stipinaite, B.S., and **Karl Reich**, Ph.D. Developmental validation of the SPERM HY-LITER™ kit for the identification of human spermatozoa in forensic samples. J. Forensic Sci., 56 (4); 853-865, 2011.

26. Jennifer Old, Ph.D., Brett A. Schweers, Ph.D., P.W. Boonlayangoor, Ph.D.[1], Brian Fischer, B.S., Kevin W.P. Miller, Ph.D., and **Karl Reich**, Ph.D. Developmental Validation of RSID™ -Semen: A Lateral Flow Immunochromatographic Strip Test for the Forensic Detection of Human Semen. J. Forensic Sci., 57(2); 489-499, 2012.

27. Alexander Sinelnikov, Anna Kalinina, Jennifer B. Old, Pravatchai W. Boonlayangoor and **Karl A. Reich** Evaluation of Rapid Stain IDentification (RSID™) Reader System for Analysis and Documentation of RSID™ Tests, *Appl. Sci.* **2013**, *3*(3), 624; Published online: 5 August 2013

28. Alexander Sinelnikov, **K.A. Reich**, Amplicon Rx™, Post-PCR Clean-up and Concentration Specifically for Forensic DNA Multiplex STR PCR Reactions, Accepted for publication, European Journal of Forensic Sciences, 09 2015

ABSTRACTS

New Approach to Statistical Reporting for Forensic DNA Analysis: Boonlayangoor, A. W., Reich, K.A., and Boonlayangoor, P.W. Published at the 14th International Symposium on Human Identification, Promega Corporation, Phoenix Arizona. September 29-October 2, 2003.

Utilizing Proficiency Testing Survey Results in Forensic DNA Laboratories.
Karl A. Reich, Liz A. Graffy and P.W. Boonlayangoor. Published at the 15th International Symposium on Human Identification, Promega Corporation, Phoenix Arizona. October 4-7, 2004.

A Novel Lateral Flow Strip Test for Rapid Identification of Human Semen (**R**apid **S**tain **Id**entification-Semen), 17th International Symposium on Human Identification, Jennifer Old, Brett A. Schweers, P.W. Boonlayangoor & Karl Reich, Nashville Tennessee October 8-12, 2006.

A Novel Lateral Flow Strip Test for Rapid Identification of Human Saliva (**R**apid **S**tain **Id**entification-Saliva), 17th International Symposium on Human Identification, Jennifer Old, Brett A. Schweers, P.W. Boonlayangoor & Karl Reich, Nashville Tennessee October 8-12, 2006.

A Novel Lateral Flow Strip Test for Rapid Identification of Human Blood (**R**apid **S**tain **Id**entification-Blood), 17th International Symposium on Human Identification, Jennifer Old, Brett A. Schweers, P.W. Boonlayangoor & Karl Reich, Nashville Tennessee October 8-12, 2006.

Developmental Validation of SPERM HY-LITER™: A Specific, Sensitive, and Confirmatory Screening Method for Human Sperm from Sexual Assault Evidence Jennifer Old, Brett A. Schweers, P.W. Boonlayangoor & Karl A. Reich 19th International Symposium on Forensic Sciences, ANZFSS Melbourne meeting – October 2008

Developmental Validation of SPERM HY-LITER™: A Specific, Sensitive and Confirmatory Screening Method for Human Sperm Detection from Sexual Assault Evidence. 19th International Symposium on Human Identification. Jennifer Old*, Brett A. Schweers*, P.W. Boonlayangoor & Karl Reich - Promega HID meeting - October 2008

Case Study: Analysis of an anorectal swab alleged to contain canine sperm using a fluorescently labeled human sperm head specific antibody. 19th International Symposium on Human Identification. Marisa Farhner, Brett A. Schweers & Karl Reich – Promega HID meeting - October 2008

Summary Results of a Blinded Study on the Effectiveness and Efficiency of using SPERM HY-LITER™ to Screen Sexual Assault Evidence for Sperm. 20th International Symposium on Human Identification. Jennifer Old Ph.D., Marisa Fahrner MS, Jie Wu Ph.D., Christian G. Westring Ph.D., P.W. Boonlayangoor Ph.D. and Karl Reich Ph.D.

Mapping Duct Tape for the Presence of Saliva Using Phadebas® Press Sheets, 23rd International Symposium on Human Identification, Lynette Johns B.S., Pravat Boonlayangoor Ph.D. and Karl A. Reich Ph.D.

Substrate Controls – A Simple Story 24rd International Symposium on Human Identification, Lynette Johns B.S., P.W. Boonlayangoor Ph.D. & Karl A. Reich Ph.D.

Solution For Partial Profiles: *Amplicon Rx*™ Post-PCR Clean-up Kit, 24rd International Symposium on Human Identification Alex Sinelnikov and Karl A. Reich

TRAINING CLASSES

Illinois Institute for Continuing Legal Education (IICLE) – Chicago IL September 2006
Faculty for IICLE DNA Evidence Course. Introduction to DNA and DNA Evidence for legal Professionals. Evidence, DNA Matching, Statistics, Defense and Prosecution Strategies, Case Review

Illinois Institute for Continuing Legal Education (IICLE) – Bloomington IL September 2006
Faculty for IICLE DNA Evidence Course. Introduction to DNA and DNA Evidence for legal Professionals. Evidence, DNA Matching, Statistics, Defense and Prosecution Strategies, Case Review

Illinois Institute for Continuing Legal Education (IICLE) – Chicago IL March 2007
Faculty for IICLE Defending Illinois Death Penalty Case – Cold Hits and Cold Cases: DNA Databases and New Technologies in Forensic DNA.

Illinois Institute for Continuing Legal Education (IICLE) – Fairview Heights IL November 2007
Faculty for IICLE Defending Illinois Death Penalty Case – Cold Hits and Cold Cases: DNA Databases and New Technologies in Forensic DNA.

Southwest Association of Forensic Sciences (SWAFS) – Austin TX October 10, 2007. Training workshop: Next Generation Sperm and Body Fluid Identification Tests: SPERM HY-LITER™ and RSID™-Saliva, Blood and Semen. Instructors: Karl Reich and Nadine Mattes.
Louisiana Association of Forensic Sciences (LAFS) – Baton Rouge LA, October 24, 2007. Training Workshop: Fluorescent Detection of Sperm from Sexual Assault Evidence. Instructors: Karl Reich and Nadine Mattes.

Northwestern Association of Forensic Scientists (NWAFS) – Salt Lake City UT –November 5, 2007. Training Workshop: Sensitive and Specific Fluorescent Detection of Human Sperm. Instructor: Karl Reich

McCrone College of Microscopy, COM700: Body Fluid Identification and Microscopic Methods of Sperm Detection for Forensic DNA/Serology/Biology. Instructor for Training class on human body fluid identification. December 11-13, 2007.

McCrone College of Microscopy, COM700: Body Fluid Identification and Microscopic Methods of Sperm Detection for Forensic DNA/Serology/Biology. Instructor for Training class on human body fluid identification. April 22-24th, 2008.

Mid-Atlantic Association of Forensic Scientists (MAAFS) – Huntington WV – April 30th, 2008. Training Workshop: Body Fluid Identification from Sexual Assault Evidence. Instructor for Training class on human body fluid identification.

McCrone College of Microscopy, COM700: Body Fluid Identification and Microscopic Methods of Sperm Detection for Forensic DNA/Serology/Biology. Instructor for Training class on human body fluid identification. August 19-21th, 2008.

Southwest Association of Forensic Sciences (SWAFS) – Little Rock AK September 25th, 2008 Training workshop: Next Generation Sperm and Body Fluid Identification Tests: SPERM HY-LITER™ and RSID™-Saliva, Blood and Semen. Instructors: Karl Reich and Ruben Nieblas.

Midwest Association of Forensic Scientists (MAFS) – Des Moines IA – September 30th, 2008. Training Workshop: Body Fluid Identification from Sexual Assault Evidence. Instructors for Training class on human body fluid identification Ruben Nieblas and Karl Reich.

McCrone College of Microscopy, COM700: Body Fluid Identification and Microscopic Methods of Sperm Detection for Forensic DNA/Serology/Biology. Instructor for Training class on human body fluid identification. December 2-4th, 2008.

SAFS/SWAFS/MAFS Combined Meeting Workshop: Current Topics and Development in Body Fluid Identification and Source Attribution. Instructor for Training Class on human body fluid identification for forensic DNA analysts. October 20th, 2009

SWAFS Workshop on Identification and Isolation of sperm from sexual assault evidence. Instructor for hands-on training class, LCM and SPERM HY-LITER™. September 23, 2010.

NEAFS/NEDIAI Combined Meeting Workshop: Forensic Body Fluid Identification Techniques – Hands-on Short Course for Saliva, Blood, Urine, Semen and Sperm. November 8th 2010.

CAC Workshop on Body Fluid Identification: Blood, Saliva Semen, Urine and Sperm. Hands-on training class, Instructors Karl Reich and Dina Mattes, Bakersfield CA, May 8th, 2012.

COURT & DEPOSITION EXPERIENCE:

Dr. Reich has been court qualified as an Expert in Forensic DNA, Forensic Biology and the interpretation of Forensic DNA Statistics in the following jurisdictions (in alphabetical order):

Dublin (Ireland)
Florida
Illinois
Indiana
Maryland
Minnesota
Missouri
New Mexico
New York
Ohio
Washington D.C.
Wisconsin

This includes cases in State and Federal courts on both criminal and civil matters. Additional details are available upon request.

Note:
Forensic DNA is defined as the methods, procedures, protocols, regulations, standards, and underlying science used to process samples, both evidentiary and reference, for obtaining genetic identity information. The collection, storage, processing, analysis of forensic evidence and the interpretation of such data are included in this definition.

Forensic Biology is defined as the methods, procedures, protocols, regulations, standards, and underlying science used to identify body fluids (blood, saliva, semen, and urine) and to identify spermatozoa from forensic evidence. The collection, storage, processing, analysis of forensic evidence and interpretation of such data are included in this definition.

Exhibit B

Affidavit: List of documents reviewed

NIRCL CODIS Report, #14, 06/05/2014

NIRCL CODIS Report, #13, 05/13/2014

NIRCL CODIS Report, #17, 12/02/2014
Electropherograms
14-00-483 DNA #14.01 A
15-00-483 DNA #14-01 B
2-00-483 DNA#29.01 AE3

NIRCL CODIS Report, #18, 01/07/2015
Electropherograms
4-00-483 DNA #01.01 B
4-00-483 DNA #01.01 G
12-00-483 DNA #1.01 SWB3
10A2-0-0-483 DNA #01.01 SWB – [swab 1, first extraction]
17-00-483 DNA #01.01 SWB2 E2

Lab notes, report 18
DNA Sample Worksheet
Allele Summary Worksheet, 1/7/2014, various
01.01 E3 wood; 29.01 A E3 fabric, 01.01 A,F wood; 01.01 B wood; 01.01 D wood; 01.01 E, I wood;
01.01 G wood; 01.01 H wood; 01.01 J wood; 01.01 K wood; 01.01 SWB3 E2; 01.01 SWB3 wood;
01.01 SWB4 wood; 01.01 SWB A2 wood; 01.01 E2 A2 wood

Allele Summary Worksheet, 1/7/15
IFI 27730-80967-Q1; 00-483 DNA #01.01 SWB E2 wood; 00-483 DNA #36-01 E2 "John Adams"; 00-483 DNA#37.01 E2 "Amarin Willford"; 00-483 DNA #38.01 E2 "Delwin L. Foxworth:

NIRCL CODIS Report, #19, 01/09/2015
Electropherograms
1-00-483 DNA #02.01 A, 10 second injection
3-00-483 DNA #02.01 C
5-00-483 DNA #02.01 E
6-00-483 DNA #02.01 F
13-00-483 DNA #20.01 A

DNA Sample Worksheet

00-483 Allele Summary Worksheet 01/09/15

02.01 A gas can; 02.01 B gas can; 02.01 C gas can; 02.01 D gas can; 02.01 E gas can; 02.01 F gas can;

20.01 A pants; 27.01 A shirt; 27.01 b SHIRT; 28.01 B sweatpants; 218.01 C E2 sweatpants; 29.01 C E2 sweatshirt

Independent Forensics, Test Report, January 22, 2014

Independent Forensics, Supplemental Test Report, June 13, 2014

EXHIBIT 10

North Chicago Police Department

Description FOLLOW UP INFORMATION - DET WARNER

SUPPLEMENTAL REPORT

Date Entered 11:48:37 AM, 04/12/2000User LUKA

Case#: 2K-002716

Narrative

SVC# 2K-002716
1-26-00 C/I MEETING
FOLLOW UP INFORMATION
INV.DET. O.WARNER, JR.

ON JAN. 24,2000 R/D WARNER MET WITH DET.L.WADE TO DISCUSS ALL INFORMATION CONTAINED IN THE REPORTED ATTEMPTED MURDER (ALLEGEDLY) OF DELWIN FOXWORTH ON 1/22/00. AT THE CONCLUSION OF THE CONVERSATION R/D ADVISED THAT SOURCES ON THE STREET WOULD BE PROBED BY R/D FOR INFORMATION RELATED TO THIS INCIDENT.

I/D L.WADE HAD SPOKEN TO A VICTIM AND WITNESS IDENTIFIED AS DELIA CONNERS. MS. CONNERS ADVISED THAT IF SHE SAW THE SUBJECT ID ONLY AS "T", SHORT, B/M. THE SUBJECT ID AS T WAS THE ONE CALLING THE SHOTS, GIVING INSTRUCTIONS. MS. CONNORS SAID SHE WOULD POSSIBLY BE ABLE TO IDENTIFY HIM. R/D WITH THE STREET NAME OF "T" AND THE DESCRIPTION AS GIVEN IN HER STATEMENT SET OUT TO FIND INFORMATION ON THE STREETS ABOUT SOMEONE FOXWORTH MAY HAVE ASSOCIATED WITH A MALE BLACK PERSON KNOWN ONLY AS "T".

MY SEARCH HAD TAKEN ME TO A PERSON WHOM AGREED TO HELP BUT DID NOT WANT TO BE IDENTIFIED OR MENTIONED IN REPORTS.

R/D ASSURED THE CONFIDENTIAL SOURCE THAT IF THE INFORMATION WAS CORRECT AND ACCURATE, THEIR IDENTITY WOULD BE PROTECTED. THE C/I DURING THIS BRIEF TIME DID NOT INQUIRE NOR DID THEY RECEIVE MONEY OR GIFTS. THE FIRST MEETING WITH THE C/I ON 1/26/00 REVEALED THE FIRST NAME AND RELATIONSHIP.

THE C/I SAID THE POSSIBLE SUBJECT RESPONSIBLE IS A CLEAN CUT WELL DRESSED HAIR CUT NEAT MED BROWN IN HIS LATE 20'S. SUPPOSEDLY THE GUY DESCRIBED IS FROM CHICAGO (SOUTH SIDE). HE HAD LOANED DELWIN MONEY AND DELWIN COULD NOT RETURN WHAT WAS LOANED TO HIM. THE C/I WAS ASKED HOW DID HE KNOW THIS INFO. THE C/I MET THE THIS PERSON THROUGH DELWIN. HE WAS INTRODUCED AS "T". THE C/I LATER FOUND THAT THE "T" WAS FOR TERRELL, BUT HE DIDN'T KNOW ANYTHING ELSE. I/D ASKED THE C/I TO FIND OUT HIS REAL NAME, DATE OF BIRTH AND ADDRESS. FOUR DAYS LATER R/D AGAIN MET WITH THE C/I. THE INFO OBTAINED FROM THE C/I IS AS FOLLOWS: TERRELL IS MARVIN TERRELL WILLIFORD, 5/5/71. HE IS A MEMBER OF A GANG UNKNOWN TO C/I. C/I HAS BEEN TO THE WESTSIDE OF CHICAGO WHERE T SPENDS ALOT OF HIS TIME. THE C/I HAS SEEN "T" GIVE ORDERS TO MANY GUYS HANG OUT UNDER A BRIDGE ON THE WEST SIDE OF CHICAGO. THERE WERE AS MANY AS 60-70 PEOPLE GATHERED. THE C/I SAID "T" DOESN'T APPEAR TO BE VIOLENT BUT AFTER SEVERAL ATTEMPTS TO SETTLE HE BECOMES HARDENED. R/D ASKED C/I TO LOCATE "T" ADDRESS. R/D NEVER HEARD FROM C/I AGAIN AFTER THAT MEETING IN NORTH CHICAGO.

Officer Signature (X) ______________________________

Case# ~~00~~-2K-002716
1-26-00 C/I meeting
Follow UP Information
Inv. Det. O. Warner Jr.

On Jan. 24 2000 R/D Warner met with
Det. L. Wade to discuss all information
contained in the Reported ~~on~~
~~of~~ (Allegedly) of Delwin Foxworth ~~Attempted~~ murder
At the conclusion of the conversation on 1/22/00.
advised that sources on the conversation R/D
be probed by R/D for information street would
to this incident.
Related
I/D L. Wade had spoken to a victim and
witness identified as Delia Conners, Ms. Conners
advised that if she saw the subject ID only
as "T", short B/m. The subject ID. as T was
the one calling the shots, giving instructions
Ms. Conners said she would possibly be able
to identify him. R/D with the street name of
"T" and the description as given in her
statement set out to find information on
the streets about someone Foxworth may
have associated with ~~named~~ a male black
person known only as "T".
My search had taken me to a person
whom agreed to help but did not want to be
identified or mentioned. 2

R/D assured the confidential source that if the information was correct and accurate, their identity would be protected. The C/I during this brief time did not inquire nor did they receive money or gifts. The first meeting with the C/I on 1/26/00 revealed the first name and relationship.

The C/I said the possiable subject responsiable is a clean cut well dressed ~~know~~ low hair cut neat med brown in his late 20's. Supposedly the guy described is from Chicago (South side). He had ~~loned~~ loaned Delwin money and Delwin could not return what was loaned to him. The C/I was asked how did he know this info. The C/I met the this person through Delwin. He was introduced as "T". The C/I later found that the "T" was for Terrell, but he didn't know any thing else. I/D asked ~~the C/I~~ to find out his real name date of birth and address. Four days later R/D again met with the C/I. The info. obtain from the C/I is as follows. Terrell is Marvin Terrell Williford 5/05/71. He is member of a gang unknown to C/I. C/I has been to the Westside of chicago where T spends alot of his time. The C/I has seen "T" give orders to many guys hang out under a bridge on the 3

16
24

//

West side of Chicago. There were as many as 60-70 people gathered. The C/I said ~~the~~ "T" dosent appear to be violent but after several attempts to settle he becomes hardened. R/D asked C/I to locate ~~his~~ "T" address. R/D never heard from C/I again after that meeting in North Chicago.

4 14 30

1

EXHIBIT 11

M/ REQ/DET WARNER
7L01AAI6Q
IL0491500
THIS NCIC INTERSTATE IDENTIFICATION INDEX RESPONSE IS THE RESULT OF YOUR INQUIRY ON NAM/WILLIFORD,MARVIN T SEX/M RAC/B DOB/19710505 PUR/C

NAME	FBI NO.	INQUIRY DATE
WILLIFORD,MARVIN T	57838NA2	2000/01/25

SEX	RACE	BIRTH DATE	HEIGHT	WEIGHT	EYES	HAIR	BIRTH PLACE	PHOTO
M	B	1971/05/05	511	165	BRO	BRO	ILLINOIS	N

FINGERPRINT CLASS	PATTERN CLASS
PO PI PI PO 19	WU WU WU WU RS LS WU WU LS LS
22 PM PI 22 20	WU

IDENTIFICATION DATA UPDATED 1991/07/16

THE CRIMINAL HISTORY RECORD IS MAINTAINED AND AVAILABLE FROM THE FOLLOWING:
FBI - FBI/57838NA2

THE RECORD(S) CAN BE OBTAINED THROUGH THE INTERSTATE IDENTIFICATION INDEX BY USING THE APPROPRIATE NCIC TRANSACTION.

END

SOS 012500 1422

STA/VALID ORIG TITLE ISS/010400

TTL/T0004864010 PURCHASED USED
VIN/1G6CD5332L4349049 PASSED VIN EDIT
90 CADI DFW 4D
ODM/ MILEAGE NOT REQUIRED
PUR/101299
LIC/PVB54 1000

OWN/WILLIFORD MARVIN T

2235 S KIKRLAND CHICAGO IL 60623
SURR TTL/T9062566009 ORIG TITLE 030399 IL
PRIOR TITLE INFO:
T8152861015 ORIG TITLE 060198 IL
T7302358016 ORIG TITLE 102997 IL
T7176792037 DUPL TITLE 062597 IL
N5973284 ORIG TITLE 021591 IL
MCO TYPE UNKNOWN

SOS 012500 1422

SOS41 DUP VINS--INQUIRE ON EACH LIC
1G6CD5332L4349049
99 J46232 CADILLAC
98 C390419 CADILLAC
~~00 F124195 CADILLAC~~
00 PVB54 CADILLAC

ARRON
Williford

W416-0196-7367.

IL01 NCIC RESPONSE
IL0491500

NO RECORD VIN/1G6CD5332L4349049

200001251529
IL0491500 OPR/KW PUR/C REQ/DET WARNER
NAM/WILLIFORD,MARVIN T SEX/M RAC/B DOB/19710505

SID/IL29253050 NAM/WILLIFORD,MARVIN T SEX/M
RAC/B DOB/19710505 HGT/511 WGT/165 HAI/BLK EYE/BRO
SKN/DRK SMT/
FPC/ FBI/
SOC/ CIR/863920 MNU/

ALIAS NAME ALIAS DOB
WILLIFORD,MARVIN TERRELL 19710505

TOTAL ARRESTS 2
CHARGES CONV OFFENSE CHARGES CONV OFFENSE
1 0 DAMAGE PROPERTY 2 0 WEAPON OFFENSES

LAST ARREST 19910322 BY ILCPD0000 CASE #
FOR UNLAWFUL USE OF WEAPONS ,FOID I D CARDS

CHF 012500 1529

AAI NO REC LEADS NAM/WILLIFORD,MARVIN T SEX/M
DOB/050571

PREVIOUS INQUIRIES WITHIN LAST 10 DAYS ON
NAM/WILLIFORD,MARVIN T.SEX/M.DOB/050571.

*** NAMES BASED ON SOUNDEX ONLY; NOT NECESSARILY AN EXACT MATCH ***

AGENCY	CDC	DATE/TIME		MSG KEY
NORTH CHICAGO PD	AAI	01/25/2000	1442	LZW5

TOTAL MATCHING INQUIRIES: 1

LZW2.AAI 00.025 14.42.39 DLN/W41659871129.

SOS 012500 1442

DL/IP STA/SUSPENDED
TDL/TIP STA/SEE ILOLNHELP
CDL/CIP STA/SEE ILOLNHELP
SCHLBUS STA/NOT A SCHOOL BUS DRIVER (SEE ILOLNHELP)

WILLIFORD MARVIN T
2238 S KIRKLAND CHICAGO 60623
SEX/M DOB/050571 HGT/5'09 WGT/165 HAI/BLK EYE/BRO
DLN/W416-5987-1129 DLC/D* DLT/DUP EXP/08051992 ISS/04101991
RES-PID CLASS/NONE
3 STOPS IN EFFECT
NO CONV LAST 12 MO
SUSP 01142000 I 13A-112B
SUSP 12301996 FINANCL
SUSP 08161993 JUDGMNT
END

CHF 012500 1442

AAI NO REC LEADS NAM/WILLIFORD,MARVIN T SEX/M
DOB/050571 DLN/W41659871129

EXHIBIT 12

LAKE COUNTY MAJOR CRIME TASK FORCE

Investigative Report

Case Number:	14-10	Type of Case:	Death Investigation	Agency:	North Chicago	Lead #:	26
Incident Date:	Jan 22, 2000	Time:		Lead Reference:	Initial Interview		
Officer:	John-Erik Anderson	Report Date:	Nov 28, 2014	Subject of Lead:	NCPD Assisting Detective		

Narrative

On 10/1/14 at approximately 11:55 a.m., Lake County Major Crime Task Force Commander George Filenko and I met with Olander J. Warner at the Lake County Courthouse Complex where Mr. Warner is currently employed as a court security officer. At the time of the original Delwyn Foxworth investigation, Mr. Warner was employed as a detective with the North Chicago Police Department. We explained to him that we wanted to discuss his recollections of the case with him. He stated he understood and agreed to speak with us.

Mr. Warner stated he only assisted on the Foxworth case and that Lawrence Wade was the lead investigator. He stated he knew Delwyn Foxworth and knew that he was a drug dealer. He stated Foxworth dealt cocaine. He also remarked that Mr. Foxworth liked young girls and at that time was interested in a 14 year old girl who was a member of the Gordon family. He stated Mr. Foxworth was advised by stay away from that particular girl by.

Mr. Warner stated just a few days after the initial incident, he was at his girlfriend's residence. He stated her name was Dee Dee. Also present was a subject named Scott Henderson. He stated he knew Mr. Henderson from North Chicago. He stated he was discussing the case with Dee Dee and at some point, Mr. Henderson interjected telling him that Mr. Foxworth owed a subject named Terrell Williford money. Mr. Henderson told him that Mr. Foxworth did not really respect Mr. Williford. He also stated Mr. Henderson told him that Mr. Williford had been in the area of Mr. Foxworth's residence a few times prior to the incident. He stated Mr. Henderson described Mr. Williford as a cool guy.

Mr. Warner told us that Mr. Henderson likes to be thought of us as a player. He liked to have information on things. I asked him if he thought Mr. Henderson had anything to do with the incident at Mr. Foxworth's residence. He stated initially he did not but is not so sure now. He stated he last spoke with Mr. Henderson about three years ago. He stated North Chicago PD only ever identified one suspect, Terrell Williford. He stated he thought Detective Wade went to the west side of Chicago to verify Terrell Williford was a gang member and if so, if he was a ranking gang member. We asked Mr. Warner why Mr. Wade would not return our calls. He stated the Wades can sometimes be moody. We asked him if he knew Keith Wade. He stated he did. He stated he thought he was possibly related to Lawrence Wade. We asked him if he knew of any association between Keith Wade and Mr. Foxworth. He stated he was not aware of any association between the two. He also told us that was not familiar with any of Mr. Foxworth's associates.

I asked Mr. Warner if I could take a sample of his DNA. He stated I could. He then signed the Consent to Collect Biological Evidence form. I took a buccal swab from the inside of Mr. Warner's right cheek. The swab was then

Prepared By: John-Erik Anderson

Approved By: [signature] 2511

LAKE COUNTY MAJOR CRIME TASK FORCE

Investigative Report

Case Number:	14-10	Type of Case:	Death Investigation	Agency:	North Chicago	Lead #:	26
Incident Date:	Jan 22, 2000	Time:		Lead Reference:	Initial Interview		
Officer:	John-Erik Anderson	Report Date:	Nov 28, 2014	Subject of Lead:	NCPD Assisting Detective		

Narrative

then sealed and subsequently entered into evidence.

Mr. Warner also related that Mr. Henderson told him that Mr. Foxworth did not respect Mr. Williford and did not think Mr. Williford was a threat to him.

Nothing further at this time.

Prepared By: John-Erik Anderson

Approved By: [signature] -2571

EXHIBIT 13

IN THE CIRCUIT COURT OF LAKE COUNTY, ILLINOIS

PEOPLE OF THE STATE OF ILLINOIS, Respondent v. MARVIN WILLIFORD, Petitioner	)	No. 00 CF 1920 Hon. Chief Judge Shanes, Presiding

AFFIDAVIT OF JENNIFER BLAGG

I, Jennifer Blagg, being first sworn and depose under penalty of perjury the following:

1. I am a licensed attorney in the State of Illinois and have conducted substantial investigation into the factual circumstances related to Marvin Williford's conviction for the murder of Delwin Foxworth.
2. As part of this extensive investigation, I spoke with a number of witnesses, including "Dee Dee," who described to me the nature of her interactions with Officer Warner and Williford. Despite being in a relationship (and then married), "Dee Dee" told me she had a sexual relationship with both Warner and Williford.
3. Dee Dee also informed me that, after Williford was convicted, Warner made a statement to the effect of "that's what he gets for having sex with my girl."
4. Even though Dee Dee was willing to speak with me at various points she was never willing to sign an affidavit and told me she faced pressure from her husband not to discuss any of these issues.
5. Officer Warner's revelation that Dee Dee was present when he met with Henderson soon after the Foxworth home invasion is not reported in any police report I have seen.

Under penalty of perjury, the undersigned certifies that the statements set forth in this instrument are true and correct to the best of my memory and belief.

JB

Jennifer Blagg Date 1/25/24

Subscribed and sworn to before me on January 25 2024.

Notary Public

OFFICIAL SEAL
MONICA FUENTES
NOTARY PUBLIC, STATE OF ILLINOIS
MY COMMISSION EXPIRES: 08/11/2027