



NATIONAL POLICING INSTITUTE
Pursuing Excellence Through Science and Innovation

Eyewitness Confidence in Showups: Results From the Lab and Field Studies



Karen L. Amendola
Laura Mickes
Maria Valdovinos Olson
Curt A Carlson
Scott Gronlund
Xueqing Chen

2025

**Confidence, Latency, and Accuracy in Eyewitness Identifications Made from Showups:
*Evidence from a Laboratory Experiment and Field Study***

**Final Research Report to the National Institute of Justice
(NIJ)**

Final Revisions December 31, 2025

Award #: 2019-75-CX-0009

Karen L. Amendola, Ph.D., *Principal Investigator*¹

Chief Behavioral Scientist

National Policing Institute

2550 S. Clark St., Suite 1130

Arlington, VA. 22202

kamendola@policinginstitute.org

(202) 833-1460 office

(571) 562-1615 mobile

ORC ID: 0000-0002-7649-5236

Award period: January 1, 2020 – December 31, 2024

Grant Amount: \$662,407 (across two studies and a national survey)

This project was supported by Award No. 2019-75-CX-0009, awarded by the National Institute of Justice (NIJ), Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the Department of Justice, the National Policing Institute, or the University of Bristol.

This report summarizes a laboratory experiment conducted in the U.K. and a field study conducted in two jurisdictions in the U.S. using archival data.

Cite as: Amendola, K. L., Mickes, L., Valdovinos Olson, M., Carlson, C., Gronlund, S., & Chen, X. (2025). Confidence, Latency, and Accuracy in Eyewitness Identifications Made from Showups: *Evidence from a Laboratory Experiment and Field Study*. Final Report by the National Policing Institute to the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice.

National Policing Institute
2550 S. Clark Street, Suite 1130,
Arlington, VA 22202
(202) 833-1460

www.policinginstitute.org

¹ Co-PIs Laura Mickes, Ph.D. & Scott Gronlund, Ph.D.

FINAL RESEARCH REPORT:

**Confidence, Latency, and Accuracy in Eyewitness Identifications Made from Showups:
*Evidence from a Laboratory Experiment and Field Study***

By:

Karen L. Amendola, Ph.D.

Laura Mickes, Ph.D.

Maria Valdovinos Olson, Ph.D.

Curt Carlson, Ph.D.

Scott Gronlund, Ph.D.

Xueqing Chen, M.S.

Acknowledgments

The authors wish to acknowledge the following individuals for their contributions to the field and laboratory studies.

With regard to the field study, we are grateful to individuals in the Hartford Police Department, State's Attorney's Office for the District of Hartford, and the Denver Police Department.

In Hartford, we are especially thankful to State's Attorney Sharmese L. Walcott and her assistant Luz Martinez who provided considerable support for this project, including on-site access to data and information necessary for the study. We also sincerely appreciate the assistance provided by former Police Chief Jason Thody and Sergeant Tyrell Jenkins in ensuring necessary police reports were made available as well. State's Attorney Walcott and Chief Thody also made referrals to individuals for conducting evidentiary strength ratings, for which we are thankful.

In the Denver Police Department, we are thankful to Chief Ron Thomas for his willing participation in this project, and to Philip Gonshak, Director of the Strategic Initiatives Bureau (SIB) who oversaw the data assembly for our project team. Others who we acknowledge for their assistance in these efforts include Daniel D. Everett, Director of the Identification Bureau; Lacinda Jaynes, Director of the Analytical Services Bureau; Andrea Webber, Records Administrator for the Department of Public Safety; and finally, Dr. Matthew Lunn, former Director of the SIB for initiating the agency's role in the project.

We are grateful for the time and efforts of Elsie Diaz-Duchesne and Dawn Gallo Scarpelli who applied their law enforcement expertise to conduct evidentiary strength ratings in a highly confidential and professional manner.

We sincerely appreciate the support and assistance provided by the NPI leadership team, including previous research Director, Travis Taniguchi, Ph.D. and especially Colby Dolly, Ph.D., Science & Innovation Director, for his guidance and ongoing support. In addition, we thank Nadia Stanley for her review of the recent eyewitness literature to update citations.

We are grateful to the National Institute of Justice and specifically two individuals: Brett Chapman, Ph.D., Senior Social Science Analyst for his scientific input and guidance throughout all phases of the project, and Cathy Girouard, Senior Grants Management Specialist, for providing important assistance as our Grant Manager for this project.

Finally, we thank Professor John Wixted, Ph.D. for his review of the penultimate version of the manuscript prior to submission to NIJ.

Table of Contents

Acknowledgments	iii
About the Authors	vi
Executive Summary	1
Overview and Problem Statement	5
Problem Statement	8
Showups	10
Are showups too suggestive?	11
Study One: Laboratory Experiment	15
Problem Statement and Literature	15
Research Questions and Hypothesis	18
Methods and Procedures	18
Participants.....	18
Actors	19
Set-Up	20
Experimental phases	20
Study phase	21
Distractor phase.....	21
Test phase.....	21
Analysis	23
Statistical Power	23
Analysis Plan.....	23
Results	23
Discussion	28
Limitations	30
Conclusions from Study One.....	30
Study Two: Field Study	31
Problem Statement and Literature	31
Research Questions and Hypotheses	33
Methods and Procedures	33
Agency sites	33
Overcoming the “Ground Truth” Problem	37
Study Protocols	43
Analysis	45
Statistical Power	45
Analysis Plan.....	45
Results	46
Discussion	50
Limitations	50
Conclusions from Study Two.....	51
Overall Conclusions	52
References	54

List of Tables and Figures

Tables

Table 1.	Participant Characteristics	19
Table 2.	Number of Responses by Response Type & Confidence Level	24

Figures

Figure 1.	Birdseye View of the Experimental Setup
Figure 2.	Receiver Operating Characteristic (ROC) Plot
Figure 3.	Confidence Accuracy Characteristic (CAC) Curve
Figure 4.	Response-Time Accuracy Characteristic (RAC) Curve
Figure 5a.	Denver Script to Instruct Witnesses
Figure 5b.	Denver Instructions for Officers
Figure 6.	Hartford Instructions to Witnesses
Figure 7.	Physical Evidence Rating Scale (from <i>Strength of Evidence Scale</i>)
Figure 8.	Confidence-Accuracy Characteristic Curve (CAC)

About the Authors

Karen L. Amendola, Ph.D. (Principal Investigator or PI)

Dr. Karen Amendola is the Chief Behavioral Scientist at the National Policing Institute. She earned her Ph.D. and M.A. in Psychology (Industrial/Organizational) from George Mason University (VA) as well as an M.A. in Human Resources Management from Webster University (MO). Amendola serves as the PI on an NIJ-funded study of organizational stress, and an Arnold Foundation-funded experiment on supervision and procedural justice. Her work focuses on a cross section of psychology and policing. Amendola has obtained, directed and/or overseen research, assessment, and training contracts and grants in excess of \$12 million (federal, state, local, and private) covering diverse areas, e.g., culture, early warning systems, ethics, hiring & promotion, officer work hours, wellness, and police-community relations. She is Chair-Elect of the Police and Public Safety Section of the American Psychological Association's Division of Psychologists in Public Service and previously served as Chair of the Division of Experimental Criminology of the American Society of Criminology (2018-2019). In 2020 - 2024, she served on the APA's Presidential Task Force on Use of Force against Black Americans, and in 2021 on the Committee on Psychological Screening for Implicit and Explicit Bias for the California Commission on Peace Officer Standards and Training (POST). She has published in esteemed journals like *Criminology and Public Policy*, *the Journals of Experimental Criminology* and *Quantitative Criminology*, *Policing: An International Journal*, and regularly serves as a peer reviewer for journals and grants. Amendola was also Associate Editor (Psychology and Law) for the *Encyclopedia of Criminology and Criminal Justice* (Bruinsma and Weisburd, Eds., 2014).

ORC ID: 0000-0002-7649-5236

Laura Mickes, Ph.D. (Co-PI)

Dr. Laura Mickes is a Professor of Psychology in the School of Psychological Science at the University of Bristol. She earned her Ph.D. from the University of California, San Diego, and has been researching recognition memory for 20 years. In the last 10 years, she has focused on eyewitness research. Professor Mickes has recently published work in *Cognitive Research: Principles and Implications* (using artificial intelligence with eyewitness memory), *Behavior Research Methods* (developed a Python toolkit to analyze and model eyewitness ID data), *Journal of Applied Research in Memory and Cognition* (confidence-accuracy characteristic analysis), *Journal of Experimental Psychology: Applied* (diagnosticity in sequential v. simultaneous lineups), and *Psychonomic Bulletin and Review* (need for better evidence in sleep studies). Grants from the US National Science Foundation, the UK Economic and Social Research Council, and the British Academy have supported her eyewitness research. Professor Mickes is the Psychonomic Society's Digital Content Editor and Consulting Editor for *Cognitive Research: Principles and Implications*. She is a Fellow of the Psychonomic Society.

ORC ID: 0000-0002-8090-9753

Maria Valdovinos Olson, Ph.D.

Dr. Valdovinos Olson is a Senior Research Associate at the National Policing Institute, where she works on a portfolio of projects in policing, corrections, and prisoner reentry. Valdovinos Olson earned her Ph.D. in Sociology from George Mason University and is an experienced researcher and project manager working on a diverse portfolio of federally-and privately- funded, multi-year research and initiatives across support, management, and leadership capacities. She has published in the *Journal of Offender Rehabilitation*, *Policing: An International Journal of Police Strategies & Management*, and *Women & Criminal Justice*, among others, and authored a chapter in the *Handbook on Prisons and Jails* (2023). Her work also includes reports for the COPS Office of the U.S. Department of Justice, including [community policing and procedural justice in jails](#) and [unmanned aircraft](#).

ORC ID: 0000-0002-5927-7296

Curt Carlson, Ph.D.

Curt Carlson, Ph.D., is a Professor of Psychology at East Texas A&M University. Professor Carlson earned his Ph.D. from the University of Oklahoma in 2008, and his research involves recognition memory and face processing, particularly focusing on eyewitness memory. This research has been published recently in journals such as the *Journal of Applied Research in Memory and Cognition*, *Cognitive Research: Principles and Implications*, *Applied Cognitive Psychology*, and *Psychology, Crime, and Law*. Dr. Carlson's research has been supported by grants from the National Institute of Justice and the American Psychology-Law Society. He serves on the editorial board for *Psychology, Crime, and Law* and the *Journal of Applied Research in Memory and Cognition*, and he is a fellow of the Psychonomic Society.

ORC ID: 0000-0002-3909-7773

Scott Gronlund, Ph.D. (Co-PI)

Scott Gronlund, Ph.D., is Professor Emeritus of Psychology and was the Roger and Sherry Teigen Presidential Professor at the University of Oklahoma. Professor Gronlund earned his Ph.D. from Indiana University and conducted research on recognition memory for over 35 years, the last 15 years focused on eyewitness research. Professor Gronlund published two review chapters on the Psychology of Learning and Motivation (2015, 2018) on system variable reforms and on the new science of eyewitness memory. Dr. Gronlund has published in the *Journal of Experimental Psychology: Learning, Memory, and Cognition*, and the *Journal of Applied Research in Memory and Cognition*, among others. Gronlund's eyewitness research was supported by grants from the National Science Foundation. Professor Gronlund served on the Editorial Board of *Psychology, Crime, and Law*, and was a Consulting Editor for *Cognitive Research: Principles and Implications*. He is a Fellow of the Association of Psychological Science (APS) and the Psychonomic Society.

Xueqing Chen, Ph.D.

Xueqing Chen recently received her Ph.D. in Cognitive & Experimental Psychology at the University of Bristol (UK), focusing on eyewitness identification accuracy and assessing eyewitness memory reliability. During her master's program, she investigated how individual differences affect identification accuracy, laying the groundwork for her doctoral research. She has presented her findings at the 2023 American Psychology-Law Society conference and the 2024 Psychonomic Society meeting. Beyond her doctoral studies, Xueqing has collaborated with memory experts in the United States, Sweden, and China. In collaboration with her supervisor and colleagues, she co-developed pyWitness, a Python package for analyzing data from eyewitness experiments, published in *Behavior Research Methods* in 2023. Through rigorous experimental design and interdisciplinary collaboration, Xueqing aims to enhance the understanding of cognitive processes in eyewitness identifications and their implications for the justice system.

ORC ID: 0000-0002-0067-0261

FINAL RESEARCH REPORT:

Confidence, Latency, and Accuracy in Eyewitness Identifications Made from Showups: A Laboratory Experiment and Field Study

By:

Karen L. Amendola, Ph.D.
Laura Mickes, Ph.D.
Maria Valdovinos Olson, Ph.D.
Curt Carlson, Ph.D.
Scott Gronlund, Ph.D.
Xueqing Chen, Ph.D. candidate

Executive Summary

Eyewitnesses' confidence surrounding identifications made from photo arrays or lineups was once believed to be unrelated to the accuracy of the IDs (see e.g., Report of the Special Master, State of New Jersey v. Henderson, 2008). However, recent evidence has reversed that view (see e.g., Katzman & Kovera, 2024; Wixted & G. Wells, 2017). This research indicates that IDs made with high confidence—*during the initial viewing/identification*—are likely to be correct. In contrast, IDs made with low confidence signal inconclusive evidence or a *chance* that the identification may be of an innocent person. This evidence also appears to hold in real-crime settings based on the results of a field study conducted in Houston (W. Wells et al., 2016; Wixted et al., 2016).

The mounting evidence on the important relationship between confidence and accuracy has led the International Association of Chiefs of Police (IACP, 2016), the National Research Council (NRC) of the National Academy of Sciences (NRC, 2014), the U.S. Department of Justice (DOJ, Yates memo, 2017), and the American Psychology Law Society of the American

Psychological Association² (APLS, 2020) to issue guidelines and/or make recommendations that law enforcement personnel record the initial³ confidence of witnesses in identification procedures as a matter of policy. However, while confidence has been demonstrated to be a strong predictor of accuracy in identifications made from photo arrays, less is known about the relationship between confidence and accuracy in field showup procedures.

A field showup is an eyewitness identification procedure that takes place shortly after a crime has occurred, when the police have found an individual who matches the description of the perpetrator. Note that lineup procedures are typically administered after some time has passed, compared to showup procedures, which occur shortly after the commission of the crime. Another major difference between showups and lineups is that showups are live, in-person procedures and lineups are photo arrays. In a showup procedure, the witness or victim is driven to a location, usually in close proximity to the crime where the detained individual is located, to undergo a showup procedure in case they are able to make an identification (see Behrman & Davey, 2001). Showups are typically conducted from the back seat of a police vehicle where the witness/victim cannot be seen by the *potential* suspect.⁴ The aim is to allow the eyewitness to evaluate whether the detained person is, might be, or is not the perpetrator.

Examining the confidence-accuracy relationship for field showups is important because field showups are a common form of eyewitness practice. The previous national survey of eyewitness identification practices by the Police Executive Research Forum (PERF, 2013)

² These guidelines are an official statement of the APLS, however, they do “not represent the position of the American Psychological Association or any of its other Divisions or subunits.”

³ At the time of the first viewing of the suspect only.

⁴ The term “potential suspect” is used to underscore the important distinction between a person who may be perceived as matching a witness’ description and a person who the police consider a suspect. Many legal and social science researchers often refer to persons temporarily detained for a showup as suspects even though the police are not likely to consider the individual a suspect until, at a minimum, a field-based identification is made or is made with moderate to high-confidence. This is reflected in many current law enforcement agency policies on showups.

revealed that field showups are conducted in over 60% of agencies, a proportion that is now just slightly higher at 64% (Amendola, et al., 2025), despite many recommendations by legal scholars, courts, and social and cognitive science researchers, as well as other guidelines that call upon law enforcement agencies, courts, and legislators to either largely avoid them or restrict their use (see e.g., Bertelsman, 2022; Brooks, 1983; G. Wells et al., 2020). Importantly, Amendola et al. (2025) found that for local law enforcement agencies, the proportion conducting showups was considerably higher (74%) than the average across state agencies, sheriffs' departments/offices, and local law enforcement.⁵ The fact that over 60% of all surveyed law enforcement agencies continue to report that they conduct showups may be evidence that most agencies are unaware of the recommendations or guidelines, or that they are unwilling to give up a potentially beneficial approach to investigating crime.⁶

Moreover, despite decades of scientific research on eyewitness identification, both judicial decisions and field-based practices have not always followed the science. This is due to many factors such as the rapid pace of scientific endeavors, inconsistent findings (often due to methodological or measurement issues), unreplicated results, anecdotal evidence, counter-intuitive conclusions, the well-known, but rarely discussed time lag between scientific consensus and field-based practice (see e.g., Stoddard, 1935), and perhaps most importantly, the more general research-practice gap (see e.g., Norman, 2010), including the faulty assumption that scientists and practitioners have the same goals, understandings, and rely upon each other for their work, or for that matter, interact in meaningful ways.

⁵ Unfortunately, this cannot be compared to PERF's findings as they did not provide a breakdown of showups by agency type.

⁶ It may also be due to considerations such as perceptions about an increasing complexity of guidance related to administration of eyewitness identification practices, or the unavailability of other leads or evidence immediately following a crime.

Scientific research on eyewitness identification has moved apace, and in recent years it could be argued that it has moved at an increasingly rapid pace. Expecting rapidly advancing science to be adopted more quickly in the field is unreasonable given evidence about the slow pace of diffusion (see e.g., Green et al., 2009; Rogers, 1962). While the length of the science-to-practice time lag has not been well-established in psychology or criminal justice, in the field of medicine, several scientists have suggested that the time lag averages about 17 years (see e.g., Morris et al., 2011) and one scientist in criminal justice, argued that it may be at least that long, if not longer in criminology (see Blomberg, 2021).

Additionally, the overall research-practice gap has a considerable bearing on the extent to which the courts or law enforcement professionals adopt scientific evidence (see e.g., Suggs, 1979). One key consideration is the fact that purely controlled laboratory (lab) research often cannot neatly fit the complexities of the field, or as Norman asserted that in practice, many variables “contradict the neat, tidy, logical assumptions of the scientist” (p. 11), consistent with concerns about external and ecological validity of laboratory-based research.

Similarly, many practitioners with lived experiences and intuitions may rely more heavily on their personal knowledge than complex, scientific studies that have not been clearly or concisely translated for their use. This perhaps has been best summarized by Green et al. (2009) who suggest there is plenty of blame to go around, noting that the gap “points variously at tradition-bound practitioners, who insist on practicing their way and believe they know their patients or populations best, and at the smugness of scientists believing that if they publish it, practitioners and the public will use it” (p. 154).

This report combines the findings from a lab study and a field study (using data from completed showups) to advance both science and practice by examining the role of confidence

(and response latency⁷ from the lab study) in predicting accuracy in showups. The goal of the project as a whole was to gain evidence from a lab experiment and field study on showup accuracy, and current law enforcement practices to better understand and potentially bridge the research-practice gap. Results from the nationally representative survey associated with this grant project are provided in a separate report to the National Institute of Justice (Amendola et al., 2025). The findings from the lab and field study reported herein provide additional evidence that moderate to high confidence identifications are associated with guilt or likely guilt.

Overview and Problem Statement

The issue of the utility of research in informing practice underlies all three research studies conducted as part of the NIJ grant described herein. As the first of two reports examining law enforcement showups (sometimes referred to as field-based identifications), this report describes both a lab study and field-based study designed to address the same questions, whether eyewitness confidence and decision latency are predictive of the accuracy of identifications. The use of both a lab and field study to examine the same basic hypotheses was seen as optimal for overcoming limitations of lab and field-based research and recognizing the relative value of both types. It was also a way to ultimately bridge the science and practice divide.

While the lab study was conducted using an experimental paradigm, the inability to manipulate certain real-world conditions in terms of ethics and practicality meant that the field study relied on previously collected (archival) data from two law enforcement agencies that captured confidence expressed by witnesses in their identifications. Unfortunately, the availability of response latency information was absent in the field study, given some limitations associated with establishing a starting point for the decision in a real-world setting.

⁷ Response latency refers to the time it takes a witness to make a decision whether it be making a positive identification, indicating that the person viewed is not the perpetrator, or unsure/unable to make a decision either way.

One of the primary concerns addressed in eyewitness research is the role that eyewitness identifications should play in identifying the actual perpetrators (the investigatory process), prosecutorial decisions (including plea bargaining), and judicial outcomes (remand or release, attribution of guilt or innocence, and sentencing). The other primary issue in eyewitness identification is the accuracy of the identifications made by witnesses.⁸

However, accuracy has been operationalized in different ways in the scientific literature, leading to considerable debate about how best to measure accuracy, a subject addressed in the literature, and in the NRC (2014) report. Previously, accuracy was assessed using what is known as a “diagnosticity ratio,” the proportion of correct identifications (IDs) as compared to the proportion of false identifications. With regard to eyewitness accuracy and its measurement, the prevailing view at this time is that the standard diagnosticity ratio (DR) falls short in determining eyewitness identification accuracy and is not based in theory. Instead, it is generally accepted that the use of “receiver operating characteristic” (ROC) analysis, a method stemming from signal detection theory provides more useful information for determining eyewitness accuracy, though it is not without its limitations (see NRC report, 2014).

Signal detection theory (SDT), or what Gronlund and Benjamin (2018) have referred to as the new-science view, facilitates the use of eyewitness evidence to help balance protecting the innocent against implicating the guilty. Confidence-based ROC analysis involves plotting the probability of accurately identifying a guilty suspect as compared to inaccurately selecting an innocent suspect, for every confidence level. A ROC curve with a large area under it shows that eyewitnesses are able to distinguish between innocent and guilty suspects (i.e., discriminability).

⁸ Either those who observed the crimes (even if only in part) or were survivors and/or victims of the crimes

The procedures that give rise to the largest area under the curve are the procedures policymakers should endorse.

Importantly, ROC analysis can separate discriminability from response bias or taking into consideration the extent to which the witness is inclined to make an identification or less inclined to do so (i.e., a liberal versus conservative bias, respectively), which a diagnosticity ratio cannot do. Thus, this new-science view demonstrates how policymakers can balance the two errors inherent to eyewitness identification decisions—false identifications of innocent suspects and the missed identifications of guilty suspects. According to signal detection theory, confidence provides information about the accuracy of the identification.

While ROC analysis measures discriminability, another type of accuracy, referred to as “reliability,” has to do with confidence. Because the IDs of suspects are of much concern to the criminal justice system, another analysis, called confidence accuracy characteristic (CAC) analysis, is used (Mickes, 2015). CAC allows scientists to plot the accuracy of suspect IDs as a function of confidence. In lab studies, there is typically a positive slope which means that confidence reflects accuracy, with higher confidence IDs being higher in accuracy than medium confidence IDs (which are also more accurate than low confidence IDs). Therefore, CAC analysis demonstrates the likely level of accuracy associated with a given level of eyewitness certainty.

In one field study, researchers measured the frequency and accuracy of rapid identifications accompanied by judgments of absolute certainty (Klobuchar et al., 2006). Twenty-six of these identifications involved strangers, and 25 of those 26 (96%) resulted in the selection of the police suspect. Although ground truth is unknown in the field, the fact that suspects were chosen at a much higher rate than matched fillers suggests that these judgments

likely exhibited high accuracy. A second field study allowed for a more detailed examination of the contribution of confidence. In this study, eyewitnesses to robberies expressed low, medium, or high confidence in their identifications. Wixted and colleagues (2016) found that the frequency of suspect identifications increased as confidence increased (from less than 20% for low confidence to over 70% for high confidence), and conversely, the frequency of filler identifications (i.e., known innocent lineup members) decreased as confidence increased. Using signal detection-based modeling, they estimated the accuracy of the high-confidence suspect identifications to be very high (97% correct). They further noted that, as confidence in suspect identifications increased, there was an increase in the proportion of cases in which independent corroborating evidence of suspect guilt was found. Both these field studies corroborate the claim of signal detection theory that confidence and accuracy are tightly coupled.

The implications of these issues and how they are addressed are substantial in terms of broader issues of justice, whether they be a) the devastating impact of wrongful convictions on innocent parties and their families, b) the inability to find and prosecute actual perpetrators leaving criminals to continue their aggressive activities, and indeed c) the effect on the criminal justice system itself in terms of perceived legitimacy of police, prosecutors, and the courts and their ability to engender trust (see, e.g., Clark et al., 2018).

Problem Statement

Research on eyewitness identification has primarily been conducted using lineups in laboratory settings. In the lab, participants typically view a video of a perpetrator committing a mock crime, and after a distractor task, are tested on their memory of the perpetrator using a lineup that contains several images of faces. Participants then attempt to identify the perpetrator.

The response options are to identify the suspect (who is innocent or guilty), a filler, or to reject the lineup (i.e., make no identification).

The general pattern of results in the last decade is that confidence is highly predictive of suspect identification accuracy under certain “pristine” conditions (see e.g., Mickes, 2015; Wixted & G. Wells, 2017; Wixted et al., 2015; Wixted et al., 2016), and perhaps even under non-pristine ones (Mickes et al., 2017). Despite confidence being demonstrated to be associated with accuracy, Berkowitz et al. (2020) argued that a “confidence trumps all” position in terms of eyewitness accuracy is overstated. However, this assertion misrepresents most of the recently published literature that simply emphasizes that there is a strong relationship between confidence and accuracy, not that it is the most important factor.

In a similar vein, much of the literature (and perhaps especially legal scholarship) tends to focus solely on wrongful convictions as being representative of most cases, despite this reflecting a salience bias. An overreliance on errors in the criminal justice system, despite their severity, may overemphasize failure (learning from error) to the detriment of learning from the many cases in which the system got it right, i.e., the process led to a rightful conviction or correct non-conviction. This does not mean that the exploration of the cases in which the system “got it wrong” is not important, as sentinel event analysis has proven informative and promising (see e.g., Friend et al., 2020; Hollway & Grunwald, 2019; McGarrell et al., 2023). However, equal attention has not been paid to the factors that contribute to the system getting it right,⁹ which may

⁹ The proportion of times the system got the right answer in criminal cases is hard to estimate as is the number of innocent parties that remain in custody. According to the Bureau of Justice Statistics (BJS), in 1990, there were 147,766 violent felony sentences in state courts alone (Langan & Dawson, 1993). By 2002 that figure jumped to 197,030 (Durose & Langan, 2004), and by 2006 it was 206,140 (Rosenmerkel et al., 2009). Absent a recent BJS report, reported rates have gone up and down since then and are likely lower. Combining data from the National Registry of Exonerations for 2013 – 2023 with data reported by Gross & Shaffer (2013) for 1989 – 2012 yields 2,036 violent crime exonerations over the 35-year period (about 58/year). Using the earlier BJS figures (approximately 183,645 violent felony convictions per year) yields a rough estimate of .032% (.000316) of the cases in which the system got it wrong or about 1 in every 3,166 cases (99.97%). Surely, wrongful convictions will continue to be uncovered, but the small proportion of them since the first DNA exoneration 35 years ago, suggest that appropriate conclusions have been reached in most cases.

be in the vast majority of cases¹⁰, and it could similarly contribute to greater evidence-based practice.

Nevertheless, while considerable research has amassed related to the above issues in lineups and photo arrays, much less examination has taken place with regard to showups (Sjöberg, 2016). Yet, showups are conducted in a majority of law enforcement agencies (Amendola et al., 2025; PERF, 2014) and little is known about the confidence accuracy relationship in these field identifications. Specifically, while some preliminary evidence from the lab suggests that both confidence and latency in showups are predictive of accuracy (see e.g., Sauerland et al., 2012, 2018), more research is needed to replicate those findings and establish a means for improving the reliability and accuracy of showup identifications.

Accordingly, the focus of the lab and field studies is to examine whether confidence and latency are associated with accuracy in showups. This research serves as a replication of lab-study findings on the strong confidence-accuracy relationship and a new field-based study using data from completed showups in two agencies.

Showups

A showup is a field-based identification of a single person who is found in the general vicinity of the crime and within a relatively short timeframe (typically within minutes to perhaps a few hours after the crime). Specifically, showups are typically used when a fairly detailed description of the perpetrator is provided by a witness/victim, when a BOLO (be on the lookout)

¹⁰ Indeed, estimating the rate of wrongful conviction is difficult (Kovera, 2024), and estimates of the proportion of wrongful convictions vary. For example, based on surveys of prisoners, some have estimated the wrongful conviction rate to be as high as 6% (see Loeffler et al., 2019), with significant variance based on crime type. In contrast, however, Cassell (2018) challenged conventional wisdom that the rate is between 1% and 5%. Using a quantitative approach, he noted that “the wrongful conviction rate appears, tentatively, to be somewhere in the range of 0.016%–0.062%.” Similarly, the first author of this report has used a different quantitative approach to come up with a rate that falls in the middle of Cassell’s estimate as described in the footnote referenced in this paragraph. And Supreme Court Justice Scalia concurred with Josh Marquis (District Attorney, Clatsop County, Oregon) who estimated an error-rate of 0.027%, despite some scholars’ criticism and characterizations as absurdly low (see e.g., Gross, 2006; Simon, 2012).

alert can be quickly distributed to patrol officers. A showup is typically less cumbersome than a live lineup or photo array both of which require live stand-ins or photo fillers. Showups require no fillers and are less time-consuming than lineups (Gronlund et al., 2012; Valentine et al., 2011).

Nevertheless, showups have often been considered to be overly or unduly “suggestive” by legal and social science scholars, courts, and advocacy and professional groups (see e.g., Bertelsman, 2022; IACP, 2016¹¹; Stovall v. Denno, 1967; United States v. Newman, 1998; G. Wells et al., 2020; G. Wells et al., 1998) primarily because the person the police suspect is purportedly obvious (Milde, 2019).

Are showups too suggestive?

While the concern over the showup procedure being too suggestive is widespread, it is possible that the problem is not the use of a showup *per se*. Instead, the procedures and practices (or lack thereof) used in many cited court cases may be to blame for the global disrepute of showups. After all, photo arrays can also be suggestive depending on the practices used in any particular case, so it is arguable (and surprising) that the showup procedure itself is considered more suggestive than other identification procedures.

Indeed, Gonzalez and colleagues (1993) asserted that showups are *not* unfairly suggestive (as compared to lineups with multiple individuals), noting that:

Although arguments about the highly suggestive nature of police showups have been expressed with great confidence by both psychological researchers and Supreme Court Justices, it is important to point out that these arguments (...) are based *entirely on intuition and anecdote* [emphasis added]. (p. 535)

¹¹ While IACP’s model policy discourages showups, it provides practice guidance for circumstances in which they are necessary..

Perhaps the presumption that showups are too suggestive stems from both an overabundance of case law in which police practices led to suggestive showups (even if they are not representative of the vast majority of showups in the field), and a lack of real-world knowledge by academic researchers and/or legal scholars as to how showups come about in many cases. If they have not spent much time in law enforcement agencies (LEAs) and/or only examined LEA-generated data, scholars may not realize how showups are actually carried out in various agencies.

In the field, after a stranger crime has been committed, many LEAs may send out a BOLO or similar alert (often immediately after a witness provides a sufficiently detailed description of the perpetrator), and the location of the crime. Generally speaking, patrol officers in the field at that time (but rarely case detectives¹²) may simply see a person potentially matching the witness' description in the general vicinity and within a relatively short time period and provide that information to their supervisors and/or dispatchers. In those cases, they are often instructed to conduct a showup. Because officers are typically assigned to respond to calls for service, they cannot focus their attention solely on searching for the suspect (unless they have no priority calls to which to respond). It is *unlikely* in these common circumstances that patrol officers have any preconceived notions about the identified person's guilt or innocence.

Moreover, how officers carry out the field showup may have a lot more to do with public perceptions of guilt or innocence of someone in temporary custody, as opposed to the use of the showup procedure itself. If an officer has temporarily detained a person but the person is handcuffed as opposed to just standing nearby, the showup is clearly more suggestive. A reasonable suspicion is all that is necessary for a temporary detainment, but to arrest or search a

¹² The availability of detectives who can get to a scene where a person matching the description is being detained is limited; because of legal issues around detention and custody, there is often not enough time for a detective to arrive on scene in the short window of temporary detention as opposed to a probable cause determination which establishes means for legal custody.

person, the police must have probable cause, which requires a higher standard or greater amount of evidence. In other words, to establish probable cause for an arrest requires more than a hunch or a witness description, e.g., the person is found carrying the stolen item, is driving the stolen car, etc. In essence, the circumstances surrounding the showup and/or the manner in which the identification procedure is conducted may render any identification practice suggestive, as opposed to it being inherent in the showup procedure.

Accordingly, it appears that the assertion and/or assumption about showup suggestibility is not inherent in the showup procedure itself, but rather based on flawed understandings, premature conclusions, salience bias, or other biases (see e.g., Heider, 1958; Taylor et al., 1979). As noted previously, one potentially flawed assumption may be that patrol officers somehow make an upfront assumption about guilt even when they do not have any prior knowledge of the case. Another is that a showup is suggestive because the police have temporarily detained a person, and as such, a witness assumes that it must be the person that the police suspect is the perpetrator. However, it could also be argued that a witness is more likely to make such an assumption when there is a live lineup or photo array, as these are typically conducted at a police station in a more formalized fashion, presumably after the police have had some time to investigate the crime. Nevertheless, the way in which a showup is conducted, rather than the procedure itself is likely the basis for much of the concern over suggestibility (and reliability).

Even if showups could be globally and accurately labeled as inherently suggestive or unreliable (that is, more so than in photo arrays), procedures and practices to minimize suggestibility and reduce bias could easily be implemented, as they have successfully been carried out in photo arrays. In fact, a quarter of a century ago, NIJ's technical working group noted that "the inherent suggestiveness of the [showup procedure] can be minimized through the

use of procedural safeguards” (NIJ Guide, 1999, p. 26). Clearly, many of the same recommendations for improving reliability and reducing bias in photo arrays and lineups can and have been applied to showups, e.g., admonitions like the perpetrator may or may not be present, or the agency will continue the investigation whether or not an identification is made.

To be sure, the NRC committee (2014) examined most issues surrounding eyewitness identification, including those associated with showups. In so doing, they did not assert that showups are too suggestive, nor did they discourage law enforcement from using them. Instead, the NRC committee recommended the development and use of standardized instructions for use with witnesses in showups¹³ and that researchers should attempt to better understand identification practices in the field (showups).¹⁴ This is not likely an oversight or limitation on the part of the committee, but may instead be an acknowledgment that the basis upon which courts, researchers, policy groups and other governing bodies have recommended limiting or severely restricting the use of showups has not been fully or appropriately established in scientific research and that LEAs will continue the practice. Even more so, the fact that the NRC committee did not suggest avoiding showups may indicate that they too believe it is the manner in which an identification practice is carried out, and not the type per se, that accounts for unreliability, inaccuracy, and misidentification. Specifically, the NRC noted that police “[should] not display the suspect in a suggestive manner (e.g., not in a police car, not handcuffed, without drawn weapons)” (p. 28).

Because courts, scholars, and others believe showups are too suggestive (see e.g., Cicchini & Easton, 2010), unreliable, and/or inaccurate (Dysart & Lindsay, 2007), they often warn against their use except in very specific, limited circumstances. While the presumption of

¹³ See Recommendation #3, p. 107

¹⁴ See Recommendation #10, p. 117

suggestibility has been widely adopted, there are limited studies on showup reliability. The photo array procedure was also once considered to be highly problematic until new research emphasized scientific and practical means for improving its reliability.

This report details the findings of a lab and field study designed to investigate the confidence-accuracy and latency-accuracy relationships in eyewitness identifications using live showups. In the lab study conducted in the U.K. with assigned perpetrators, and a U.S.-based field study, using tools to approximate ground truth in two jurisdictions, served to strengthen the demonstrated importance of confidence in predicting accuracy in showups.

Study One: Laboratory Experiment

For the most part, scientists examining showups have relied upon individual images presented to participants. In those studies, confidence has not necessarily been a good predictor of accuracy compared to lineups, because lineup research involves showing images of the suspect among several individuals who are matched to the suspect's description and are known to be innocent. In this study, however, we conducted an experiment to determine if confidence and response time indicate accuracy in live showups.

Problem Statement and Literature

The focus of research on eyewitness identification procedures has largely been on one type of accuracy: discriminability. Discriminability is the ability to distinguish an innocent suspect from a guilty one. In discriminability experiments, showups underperform lineups. Specifically, participants do not discriminate as well when memory is tested on a showup compared to a lineup (Key et al., 2015; Gronlund et al., 2012; Mickes, 2015; Wetmore et al., 2015; Yarmey et al., 1996).

A second type of accuracy in eyewitness identification procedures is reliability. Reliability is the likelihood that the identified suspect is the true perpetrator (Mickes, 2015). To measure reliability, the proportion correct at each level of confidence is plotted in a confidence-accuracy characteristic (CAC) plot. If found to be reliable, suspect identifications made with high confidence are higher in accuracy than medium confidence identifications, and the medium confidence identifications are higher in accuracy than low confidence identifications. Establishing reliability is an essential element in ensuring accuracy in eyewitness identification.

The procedure for studying showups in the lab is the same as for lineup studies, except that only one image of a face is viewed. This image is either of the person who was shown in the video perpetrating a mock crime (the assigned perpetrator) or another person (the innocent suspect). Participants are asked to report whether the person in the photo is the same person shown in the video (“old”) or not (“new”) and then rate their confidence in their decision. The pattern of results from lab-based showup studies using images differs from that found when memory is tested on lineups. With image-based (as opposed to live) showups, high confidence is not as high in accuracy as it is with photo lineups (Carlson et al., 2021; Colloff & Wixted, 2020; Key et al., 2015; Mickes, 2015; Sauerland et al., 2018; Wilson et al., 2018).

Diagnostic-feature-detection theory (DFD) provides an explanation for why confidence is not as predictive in a showup as it is in a lineup (Wixted & Mickes, 2014). The reason is that seeing only one person, as opposed to a lineup containing multiple persons, makes it more difficult to distinguish between various features available when multiple faces are shown (see Colloff & Wixted, 2020).

In real life, however, showups are in-person identification procedures, and little research has been conducted on how law enforcement agencies administer them in the field. Though there

have been a few live showup experiments, the relationship between confidence and accuracy under these conditions is still not fully understood (e.g., Valentine et al., 2011). Findings from live versus simulated field showups are mixed. For example, Eisen et al. (2017) found that accuracy actually decreased as confidence increased, a pattern that is the opposite of typical findings where accuracy increases as confidence does and Eisen et al., 2022 also found participants to be confident in 50% of false identifications. However, these researchers had the field suspect standing next to two uniformed police officers with his hands behind his back as if he were handcuffed. This practice made the showup more suggestive. It is not surprising, therefore, that confidence was more meaningful when participants were tested in the lab on live showups. Under these conditions, only 27% of the false identifications were made with high confidence. These results suggest that confidence is less predictive of accuracy when memory is tested on a live showup (in a lab setting) and may not differ from static showups. Nevertheless, data from these studies suggest that showup confidence may be inflated in false identifications, even though it may be related to accuracy. This underscores the need for more research to disentangle discrepant findings.

In addition to confidence being a useful indicator of accuracy (at least in lineups), response time, or the time it takes to make an identification from a lineup, appears to also be indicative of accuracy. Specifically, faster identifications are associated with higher accuracy in the identification than slower identifications (e.g., Seale-Carlisle et al., 2019). Eisen et al. (2022) also measured response times during live showups and found that accurate participants were faster in making their identifications than inaccurate participants. Sauerland et al. (2018) measured confidence and response time using showup images and found that high confidence

identifications made quickly were highly accurate, more so than lower confidence identifications that were made quickly and high confidence identifications that were made slowly.

Research Questions and Hypothesis

In this study, we set out to answer the following two questions:

1. Does the information about witness confidence in identifications made from a showup procedure increase the probative value of eyewitness evidence?
2. Does the time it takes to make an identification from a showup procedure increase the probative value of eyewitness evidence?

Based on findings from a pilot study¹⁵, we hypothesized that confidence and response time would predict accuracy in the identifications.

Methods and Procedures

Participants

Participants ($N = 229$) were recruited at the University of Bristol Neuroscience Festival in early March 2023 and were randomly assigned to a target-present showup ($n = 120$) or a target-absent showup ($n = 109$). The participants were, on average, 28 years old ($SD = 14.57$; range 8-73 years). Anyone over 8 years old could participate in the experiment, but for minor participants under 16 years, signed consent was obtained from their legal guardians. Minor participants ($N = 18$, age in years: $M = 12$; $SD = 1.97$) were accompanied by their guardians throughout the experiment.¹⁶ Table 1 details the participant characteristics.

¹⁵ The pilot study was begun prior to the publications of the live showup experiments (Eisen et al., 2017; Eisen et al., 2022).

¹⁶ Ethical approval was granted by the School of Psychological Science Research Ethics Committee; reference number 13830.

Table 1

Participant Characteristics (N = 229)

	<i>N</i>	%
Gender		
Female	154	67
Male	72	31
Unknown	3	1
Race/Ethnicity ^a		
White	180	79
Asian	27	12
Black	4	2
Hispanic	4	2
Other	13	6

Note. Totals may not add up to 100 because of rounding.

^aThree participants did not provide their ethnicity.

Actors

Ten actors served as the suspects. Four Asian students (2 female and 2 male) volunteered as research assistants and were assigned as actors. Six white students (3 female and 3 male) responded to an advertisement email for actors and passed the screening. All actors were young adults with similar features. Male actors had to meet the following requirements: 1) short brown or dark hair; 2) height ranging from about 5 ft, 6½ in. to 6 ft (172–183 cm); 3) weight ranging from about 165–185 lb (75–84 kilograms); 4) light or brown eye color; 5) shaved with no beards, mustaches, etc.; and 6) no distinctive features, i.e., no tattoos, piercings, or scars. For female actors, they had to meet the following requirements: 1) medium to long, brown or dark hair; 2)

height ranging from about 5 ft, 2½ in. to 5 ft 5 in. (160–168 cm); 3) weight ranging from about 110 –141 lb (50–64 kilograms); 4) light or brown eye color; 5) no distinctive features, i.e., no tattoos, piercings, or scars.

To test memory for the person, not the clothing, all actors wore the same black Bristol Neuroscience Festival T-shirt and a black hat for the study phase and wore their own plain hoodies without a hat for the test phase. During the entire experiment, the male actors wore blue jeans, and the female actors wore black leggings. All the actors wore white sneakers and shoe covers.

Set-Up

As shown in **Figure 1**, a makeshift room made of movable partitions was assembled to house the actors, so they were not visible to the festival's attendees. The marked spots are where the actors stood. Participants observed the actors while being 12 feet away from the study and test spots. For the target-present showups, the same actor appeared at both the study and test spots (i.e., the suspect is the guilty perpetrator). For the target-absent showups, actors of the same gender and race were randomly paired. One of the paired actors appeared during the study phase, and the other appeared during the test phase (e.g., the suspect is innocent). We followed a counterbalanced actor list to ensure an equal number of target-present and target-absent showups. The actors were assigned as targets an equal number of times.

Experimental phases

In eyewitness identification studies, there are typically three phases: a study phase, a distractor phase, and a test phase. The study phase is analogous to an eyewitness encoding the crime. The distractor phase is similar to the break between the commission of the crime and the identification procedure. The test phase is the identification procedure where the eyewitness

attempts to identify the perpetrator.

Study phase

After completing the informed consent process (see the blue icon labeled 1 in **Figure 1**), participants were led by a research assistant to stand at the study-marked spot (blue icon labeled 2). Adult participants were told, “*You are about to see a person for many seconds. Please try to remember what they look like.*” To make it slightly more engaging to minor participants, they were told, “*A ‘baddie’ will appear at that spot for a few seconds. Please try your best to remember their face, but not their clothes or shoes.*” After the research assistant ensured the participants understood the instructions, an actor playing the perpetrator stood at the study spot and silently counted to ten. The actor looked straight ahead, facing the witness participants with a neutral facial expression and without making prolonged direct eye contact.

Distractor phase

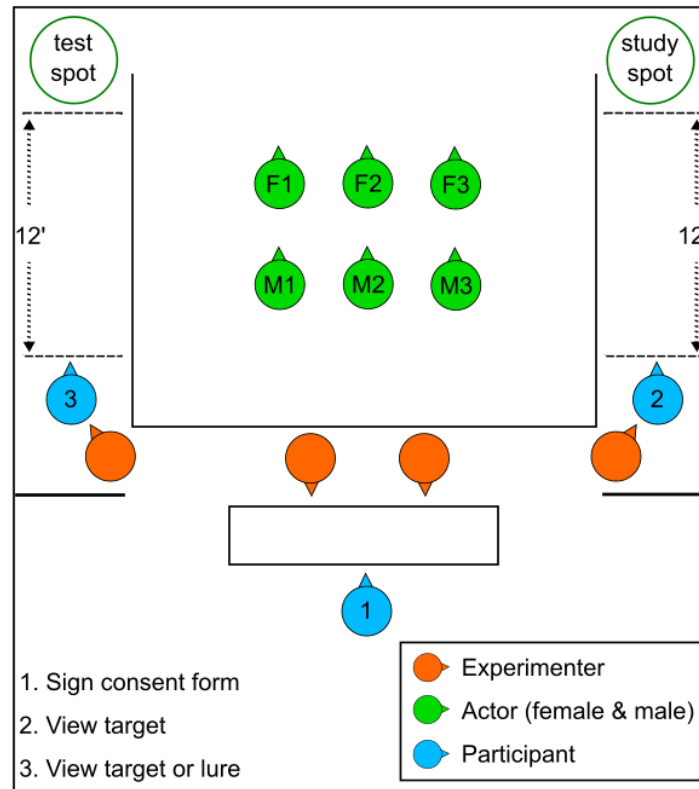
In eyewitness identification studies, there is typically a distractor phase between the study and test phases. This breaks up the study session into its own separate episode and reduces the likelihood that participants will use working memory instead of long-term memory. During the distractor phase, participants were asked to provide their own demographic information and engaged in a 5-minute activity. Then, research assistants shared neuroscience knowledge about memory-related brain regions with participants and chatted about their experiences at the festival. Minor participants also played a game of reciting the alphabet backward.

Test phase

Participants were then led to the test-marked spot (blue icon labeled 3 in **Figure 1**) and were given additional instructions based on their age group. Adult participants were informed, “*You are about to see a person. The person you saw earlier may or may not be the*

Figure 1

Birdseye view of the experimental setup



person you are about to see. If the person is the same as earlier, please circle 'yes' on your answer paper. Otherwise, circle 'no'. Minor participants were told, “You will see a person showing at that spot. They may or may not be the baddie you saw before. If they are, circle 'yes' on your answer paper. Otherwise, circle 'no'.”

After the research assistant ensured that the participant understood the instructions, the suspect, who was either the same (target) or a different actor (lure), stood at the test spot, looking straight ahead with a neutral expression to ensure participants could clearly observe their face. The research assistant started the timer once the suspect showed up and stopped it once the participants made their decision. Participants then rated their confidence level in their decision on a scale of 3 levels: definitely, probably, and guess. Afterward, they were debriefed on the

study and their participation.

Analysis

Statistical Power

With information gleaned from the Houston Field Study, we determined that we needed 216 cases with choosers to establish sufficient statistical power for our analysis. With 229 participants, and two individuals that guessed, we obtained sufficient power for our analyses.

Analysis Plan

The analyses of interest are confidence-accuracy characteristic (CAC) and response-time accuracy characteristic (RAC) (Mickes, 2015; Seale-Carlisle et al., 2019). Both CAC and RAC are measures of reliability because they provide estimates of the likelihood that the suspect identified from the showup is guilty. Moreover, both CAC and RAC curves reflect ID accuracy as a function of confidence level or response time, respectively. If confidence is not predictive of accuracy, we do not expect to see a slope, and if confidence is predictive of accuracy, we expect to see a positive slope. Likewise, if response time is not predictive of accuracy, we do not expect to see a slope, and if response time is predictive of accuracy, we expect to see a negative slope. All analyses were conducted using pyWitness (Mickes et al., 2023).¹⁷

Results

The overall correct ID rate = 0.856, false ID rate = 0.064, and $d' = 2.583$. The participants first indicated if the suspect at the test spot was the person shown at the study spot or “old” meaning seen before (the perpetrator), or a new person and then rated their level of confidence in their decision. The scale including the guesses covered six points: definitely new, probably new, guess new, guess old, probably old, definitely old. The confidence scale while also considering

¹⁷ The data and code are available on OSF (<https://osf.io/9rhkv/>).

guesses, was as follows:

"Definitely new" means participants believe the suspect is NOT the perpetrator.

"Probably new" means participants think the suspect may NOT be the perpetrator.

"Probably old" means participants think the suspect may be the perpetrator.

"Definitely old" means participants believe that the suspect is the perpetrator.

Table 2 shows the number of identifications by response type for target-absent and target-present lineups for each level of confidence. Only two guess responses were made and accordingly were collapsed into those categories (1 “guess new” combined with “probably new” and 1 “guess old” combined with “probably old”), leaving a 4-point scale for the analyses.

Table 2

Responses by response type for target-absent and target-present showups by confidence level

Lineup	Response Type	Confidence					
		definitely new	probably new	guess new	guess old	probably old	definitely old
Target-absent	Reject ID	85	17	0	0	0	0
	Suspect ID	0	0	0	0	7	0
Target-present	Reject ID	7	9	1	0	0	0
	Suspect ID	0	0	0	1	28	72

Discriminability was very good ($d' = 2.58$; area under the curve = 0.95), as shown in the receiver operating characteristic (ROC) curve in **Figure 2**.

Since both CAC and RAC are graphical measures of reliability that provide estimates of the likelihood that the suspect identified from the showup was guilty, we analyzed the data using CAC to test our first hypothesis that confidence will provide predictive information about the accuracy of the identification.

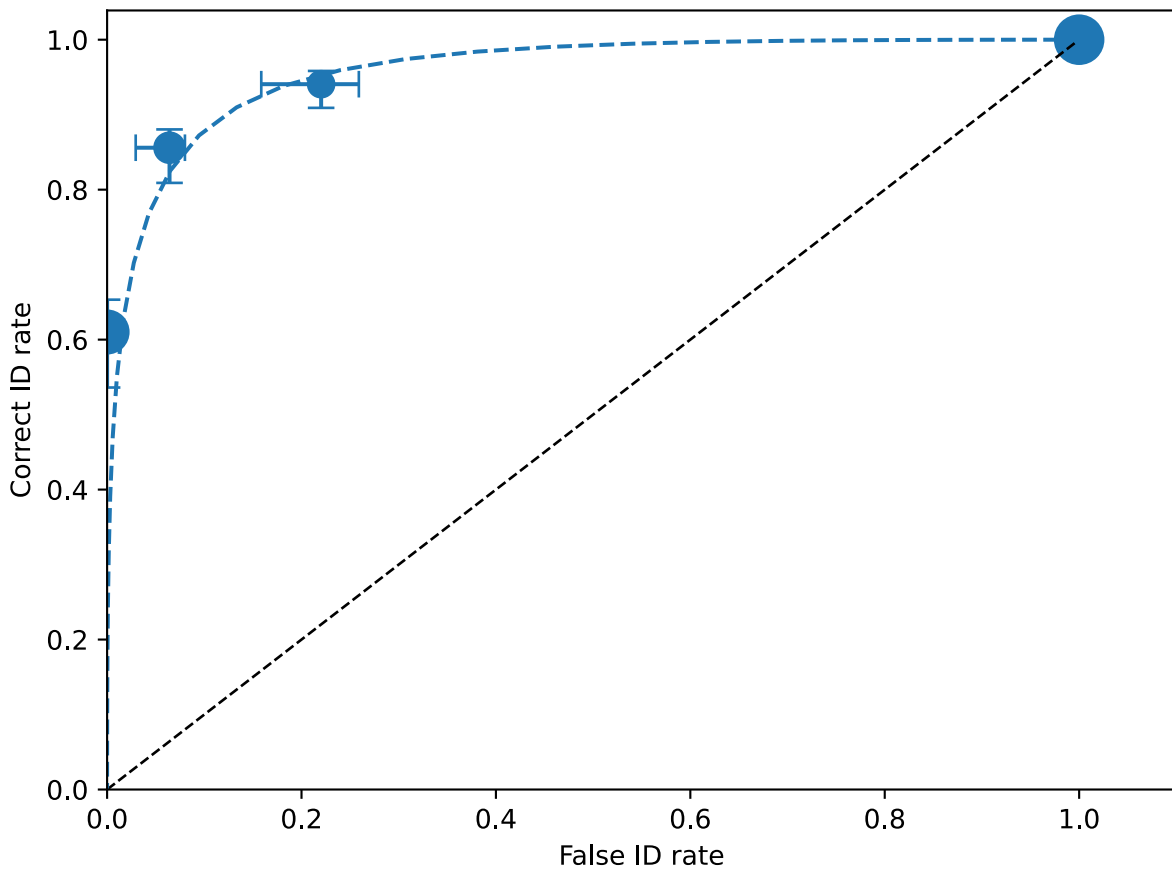
CAC is given by

$$CAC = \frac{CID_{conf}}{CID_{conf} + FID_{conf}}$$

where CID_{conf} is the number of correct suspect identifications made with a level of confidence from target-present showups, FID_{conf} is the number of false identifications made with a level of confidence from target-absent showups. CAC is computed for each binned level of confidence (definitely new, probably new, probably old, definitely old), and plotted in a CAC plot.

Figure 2

Receiver operating characteristic (ROC) plot

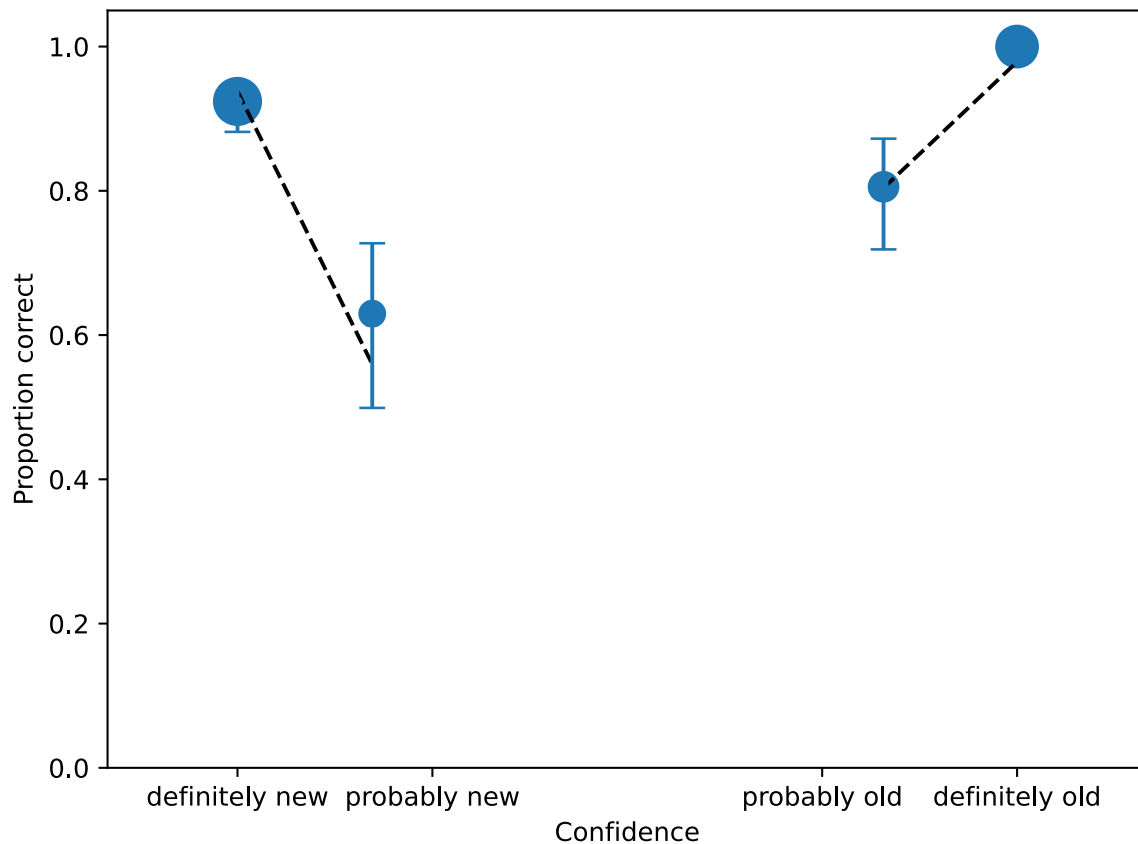


Note. The error bars are 68% confidence intervals based on 200 bootstraps. The blue dashed line is the equal variance signal detection model fit. The point sizes represent relative frequency, and the black dashed line signifies chance performance.

Figure 3 shows the CAC curve for Study One. The highest confidence responses are higher in accuracy than the lower confidence responses. The *definitely old* responses were higher than the *definitely new* responses. This finding is consistent with the literature (e.g., Tekin et al., 2021; Mickes et al., 2011).

Figure 3

Confidence accuracy characteristic (CAC) curve for Study One



Note. The error bars are 68% confidence intervals based on 200 bootstraps. The point sizes represent relative frequency.

We analyzed the data using RAC to test our second hypothesis, which is that response time will provide predictive information about the accuracy of the identification.

RAC is given by:

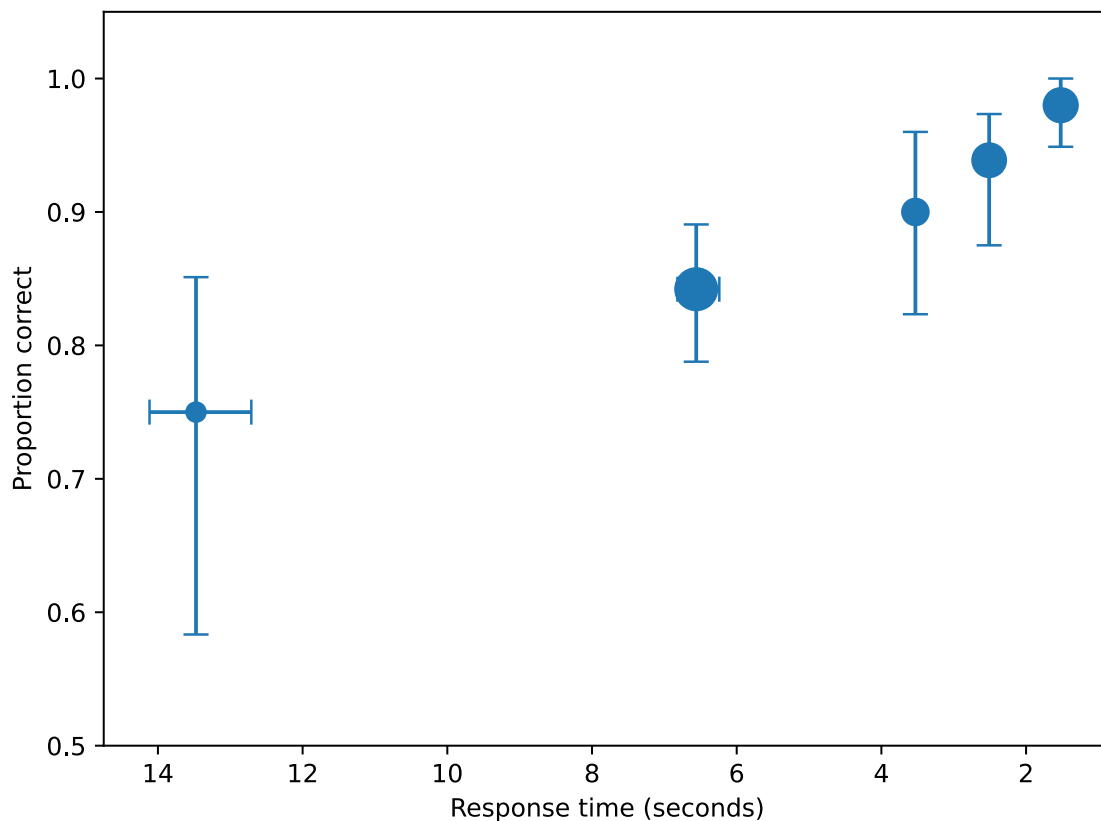
$$RAC = \frac{CID_{rt}}{CID_{rt} + FID_{rt}}$$

where rt is response time, CID_{rt} is the number of correct suspect identifications made with a certain speed from target-present showups, FID_{rt} is the number of false identifications made with a certain speed from target-absent showups. RAC is computed for each binned level of response time (from slow to fast) and plotted in an RAC plot.

Figure 4 shows the RAC curve. It shows that faster responses were more accurate than slower responses. Broadly, the pattern is linear with response time, suggesting that as responses become quicker, they become more accurate.

Figure 4

Response-time accuracy characteristic (RAC) curve



Note. The error bars are 68% confidence intervals based on 200 bootstraps. The point sizes represent relative frequency.

Discussion

In this study, we tested the hypothesis that confidence and response time would predict accuracy in live showups. Indeed, our experiment testing memory using live showups yielded impressive indications of accuracy. Confidence is often an indicator of accuracy when memory is tested on lineups but is typically less informative in showups using images. We found that confidence is reflective of accuracy in live showups. Specifically, identifications from live showups made with high confidence were more accurate than those made with lower confidence. We also found that response time is another predictor of accuracy. Identifications that were made quickly were more accurate than those made more slowly.

Live and image showups reveal different findings about confidence because in most showup experiments researchers use images instead of live people. In these experiments, we have found that confidence is less predictive of accuracy compared to lineups. A theoretical explanation derives from the DFD theory (Wixted & Mickes, 2014). As in showups, when there are no other faces to help witnesses focus more on diagnostic features and less on nondiagnostic features, performance suffers (i.e., lower discriminability, lower CAC profile). However, in our experiment, the CAC plots resemble CAC plots from fair lineups. It is possible that because the busy venue where we collected data provided some unmeasurable diagnostic information, participants were able to use the faces of the event's attendees as comparison faces. Another explanation may simply be that external validity is greater yielding more precise and accurate findings representative of those in real-world situations.

It is also possible that participants in our experiment benefitted more from the encoding specificity principle (Tulving & Thompson, 1973) than from static showups. The encoding

specificity principle holds that memory is improved when the stimuli presented during encoding are similar to those presented during retrieval phases. We believe that the encoding specificity principle may be at play. When images are used, participants typically view a mock crime incident and are then shown an image of a face that may be of an innocent or guilty suspect. In our experiment, with live lineups, the stimuli (i.e., the actors) were shown standing at the same distance from the participants during encoding and retrieval (though they moved to a different location for the test phase; see **Figure 1**).

Another potential reason why live showups yield better results than static images may be that they are richer in stimuli than static images of faces. Wetmore et al. (2015) suggest that in-person showups allow for more retrieval cues and, therefore, predicted that witness performance could be better with a live in-person showup compared to a static image showup. We attempted to minimize bias using memory for clothing or location by having the actors change clothes during the storage phase and moving the test location to a different spot from the encoding location (see **Figure 1**).

As in lineup studies (e.g., Brewer et al., 2006; Dunning & Perretta, 2002; Holdstock et al., 2022; Seale-Carlisle et al., 2019), quick identifications were more accurate than slower ones. Our findings align with those of other showup experiments, including static showups (Sauerland et al., 2018) and live showups (Eisen et al., 2022). If this finding continues to be replicated, it indicates that the police should collect response time during showup procedures. However, it should be noted in our field study that identifying a specific time when a witness should view a suspect when being driven can be complicated by the person attempting to say something as they are approaching the location where the suspect is being temporarily detained, as was explained to us by at least two agencies (one of which decided not to participate).

Limitations

To keep participants engaged in the experiment, we made the distraction task short (5 minutes). Future experiments should measure reliability (using CAC and RAC) in live lineups with longer retention intervals, similar to real witnesses' retention intervals during police proceedings. Another limitation may be that participants knew this was an experiment, and thus, their identification was of no consequence to the identified suspect. Because study participants are calm, they likely encounter little arousal or stress that could interfere with memory tasks like encoding, storage, and retrieval. Therefore, there is a need to conduct research with police partners in the field, using real eyewitnesses to real crimes. This is why this project included both a lab and field study.

Conclusions from Study One

We investigated the reliability of live showup identifications in a controlled, lab experiment. Our findings revealed that identifications from live showups can be reliable when considering confidence and response time. Our experiment contributes to a small but growing set of studies investigating live showups. These findings have important applied implications. While showups are believed by many to be highly suggestive, they are a common tool used by law enforcement, as identified in the PERF Report (2013) and in the National Policing Institute's nationwide survey (Amendola et al., 2025). Knowing that live showups have indicators of accuracy should provide some reassurance in their continued use. If this pattern holds across a range of experiments, identifications stemming from showup procedures may be more reliable than has been suggested, despite continued arguments as to their suggestibility.

Study Two: Field Study

In Study One, we demonstrated a strong confidence-accuracy relationship (as well as one between latency and accuracy), yet there is an absence of research on showup confidence conducted in law enforcement agencies. This is due to at least two factors, first, there are not that many law enforcement agencies that have reported capturing confidence from witnesses during showup procedures, and identifying those agencies is difficult. The other reason is that in the real world, there is the ground truth problem, i.e., that rarely is there proof positive that the identified suspect is indeed the perpetrator of the crime, even in the rarest cases where DNA evidence is available.

Problem Statement and Literature

Despite their frequency of use, showups have been deemed by many scholars and practitioners to be too suggestive as previously indicated. Because of this concern, the current IACP policy (among others) recommends avoiding showups when photo arrays can be used instead (IACP 2016). While some have recommended that agencies abandon the use of showups altogether, the extent to which agencies in the U.S. have done so was examined in the nationwide survey as part of this grant project, with about 74% of local law enforcement agencies, and 64% of agencies overall¹⁸ reporting the use of showups (Amendola et al., 2025), a very small increase of about 2% over the prior study conducted by PERF (2014). Clearly, a majority of law enforcement agencies are unaware of that guidance and/or simply do not plan to abandon a practice that sometimes proves fruitful in investigating and prosecuting crime.

Although field researchers had not yet sufficiently examined the confidence-accuracy

¹⁸ Across local, state, and sheriff's agencies

relationship in live showups, laboratory-based data previously suggested this relationship may be less indicative of accuracy, with some demonstrating that confidence in those identifications is inflated (Eisen et al., 2017). However, in Study One, our results demonstrated a strong confidence-accuracy relationship.

It is unclear what may account for a weaker relationship between confidence and accuracy in showups versus photo arrays, but some have posited it is due to an inability to make comparisons across facial features that reduces this relationship (Wixted & Mickes, 2014). Notably, the studies to date have relied primarily on images of suspects rather than live in-person suspects as in real showup procedures. These studies have shown that high confidence IDs from photo showups are lower than expected (Mickes, 2015; Wetmore et al., 2015). However, in a recent lab-based pilot experiment with live showups (using actors), confidence in a showup decision was, in fact, informative of accuracy (Wilson et al., 2018) and was also replicated in Study One reported herein. The findings from lab studies have been more conclusive than those from the field since there are far fewer published field studies. Accordingly, the field test in Study Two is designed to explore this issue more systematically. The reliability and diagnosticity of live showups in the field have yet to receive sufficient examination.

Additionally, evidence has routinely demonstrated that the time it takes a witness to identify a suspect (also known as latency) is probative; the faster those decisions are made, the more likely they are to reflect ground truth (see e.g., Sauerland et al., 2012; Sauerland & Sporer, 2009). Information about the latency of eyewitness decisions in showups from the field has the potential to shed light on the extent to which it is predictive of accuracy in these circumstances, as well as to help identify ways to make field showups more reliable.

Given the frequency of showup procedures conducted in the field¹⁹ and experts' assessments that this practice is "likely to continue" (Neuschatz et al., 2016, p. 66), these questions about confidence, latency, and their relationships to accuracy are of paramount importance. This is especially the case because courts, scholars, and other groups have addressed the admissibility of showups and/or advised against them as reported previously. Further, eyewitness science has moved apace over the past decade resulting in an upending of some prior research and greater consensus on many issues in the eyewitness identification field (Gronlund & Benjamin, 2018). Hence, it is necessary to obtain a better understanding of showups as they are conducted in the field.

Research Questions and Hypotheses

In this field study, we attempted to address the same two important questions as were employed in Study One. Specifically, we were interested in answering the following questions:

1. Does the information about confidence in identifications from live showups increase the probative value of eyewitness evidence?
2. Does the time it takes to make an identification from a live showup procedure (response latency) increase the probative value of eyewitness evidence?

We hypothesized that both confidence and response latency would provide useful information about the accuracy of eyewitness identification.

Methods and Procedures

Agency sites

In this field study, we were able to identify two jurisdictions that have been collecting

¹⁹ PERF's national survey was based on 2011 data (PERF 2013; the forthcoming National Policing Institute (NPI) survey was based on data collected in 2021 (Amendola et al., forthcoming 2025).

confidence in showups for several years²⁰ to potentially serve as research sites: the City and County of Denver, Colorado, and Hartford, Connecticut (police department and prosecutor's office). These jurisdictions were selected from agencies in our nationwide survey that reported that they collected confidence in showups and were large enough to generate a sufficient number of cases with showups, were willing to participate, and had the available data necessary for the evaluation. To be specific, many law enforcement agencies in the U.S. that conduct showups rely solely on a binary judgment, i.e., "is this the perpetrator" allowing for just a yes or no decision. This forced choice format, sometimes recommended by prosecutors, does not leave room for unsure responses and may in fact push a witness to employ a liberal bias (say "yes" even when unsure) or conservative bias (say "no" for fear of making an incorrect ID). In essence, it can render confidence statements unnecessary or not useful. One of the key reasons we chose to work in these two jurisdictions, was that these agencies had protocols for minimizing bias and allowing for confidence to be captured.

In both jurisdictions, the police departments provided data, and in Hartford, data was also accessed from and reviewed on-site in the State's Attorney's Office. To preserve real-world conditions, we relied on archival data (i.e., data that had already been collected in these police departments), thereby not manipulating how the real world showups had been conducted.²¹

In both jurisdictions, in-place protocols addressed prior concerns over bias in eyewitness

²⁰ This is because we conducted, as part of this grant, a nationwide survey of agencies and were thus able to identify many (but a very small proportion of) agencies that responded that they do collect such data. Some were from agencies that were too small to have enough accumulated data to be able to examine these cases, and others who we contacted who were not able to participate.

²¹ It would be premature and inappropriate to conduct an experiment in the cooperating agencies by changing current eyewitness practices and procedures over a long period without sufficient knowledge of how such changes may impact actual criminal cases. It was necessary to obtain buy-in by from law enforcement and prosecutors, and to abide by state laws or state attorneys' general guidelines, if any. Consequently, we made it a point to not disrupt existing police practices by manipulating an independent variable, or to request that agencies change instructions or require them to follow an unfamiliar set of protocols, as the demands of research should not result in significant field disruptions or risk exposure to their law enforcement partners. In addition, given the extensive time some cases take to reach a disposition (through plea bargains or adjudication), there was not enough time in the original (or extended grant timeline) to collect a sufficient number of cases for statistical power while awaiting cases to reach final dispositions, especially in light of serious time gaps in the research due to Covid-19, which delayed the project substantially.

procedures. In Denver, the use of an “advisement card” provides instructions to officers and witnesses. The advisement card that officers verbally present to the witness in Denver is provided in **Figure 5a**, and the instructions to the officers are provided in **Figure 5b**.

Figure 5a

Denver lineup or showup script for advising witness

EYEWITNESS IDENTIFICATION ADVISEMENT CARD

Officer delivering advisement will follow the script below:

“In a moment I am going to ask you to view someone/some people. The person(s) who committed the crime may or may not be one of the people you are about to view.

You should not feel you have to make an identification. It is just as important to clear innocent people from suspicion as it is to identify the suspect. The investigation will continue whether or not you make an identification. If you do select someone, please do not ask me about that person, as I am unable to share any information at this stage of the investigation.

Because you are involved in an ongoing investigation, to prevent compromising the investigation, you should avoid discussing this identification procedure or its results. Keeping the results of this procedure to yourself is necessary to preserve the integrity of the procedure.

Do you understand the instructions I have given you? Do you have any questions before we begin?”

Figure 5b

Denver instructions to officer(s) administering lineup or showup

Officer Instructions:

1. Remember to remain neutral. Do not comment on the identification before, during, or after the show-up.
2. Remember to turn on your body-worn camera prior to beginning the show-up.
3. During the show-up, stand out of the victim/witness' line of sight, where practical, but still observe and hear the victim/witness.
4. If victim/witness makes an identification, ask their level of certainty
 - A. Confident
 - B. Somewhat confident
 - C. Not confident
5. Language Line: 1-844-814-1941. Interpreters for the hearing impaired may be requested through the Denver 911.

Similarly, in Hartford, the agency provides officers a single card with witness instructions for photo arrays, live lineups, and showups, see **Figure 6**.

Figure 6

Hartford instructions to witnesses

1. Please listen carefully as these instructions are read aloud to you. Each one of the instructions is equally important. You have been given a copy of these instructions to read along with the officer if you wish. After you have read the instructions, please initial after the instruction to indicate that you have read the instruction and understand it. (Initial _____)
2. You will be asked to view an array of photographs, a person or group of persons, and each photograph or person will be presented one at a time; (Initial _____)
3. It is just as important to exclude innocent persons as it is to identify the perpetrator; (Initial _____)
4. The person(s) in the photographic lineup, live lineup or show-up may not look exactly as they did on the date and time of the offense because features like clothing, facial or head hair can change; (Initial _____)
5. The perpetrator may or may not be among the person(s) in the photographic lineup, live lineup or show-up; (Initial _____)
6. You should not feel that you must make an identification; (Initial _____)
7. You should take as much time as needed in making a decision; (Initial _____)
8. If you are able to make an identification of someone, you will be asked to describe in your own words how certain you are of that identification; (Initial _____)
9. (N/A in show-up) Even if you are able make an identification, you will be asked to finish the procedure by looking at all of the photographs or all of the individuals until you have completed looking at each one; (Initial _____)
10. If there are other witnesses, you must not indicate to them that you have or have not made an identification of a person; (Initial _____)
11. (N/A in show-up) The officer administering this procedure either does not know whether any of the people in the photographic array or in the lineup were involved in the crime or does not know the order in which you are viewing the photographs; (Initial _____)
12. If you do select someone, the officer will not be able to provide you with any information about the person you have selected; (Initial _____)
13. If you select a person or photograph you will be asked to provide a statement about this process and the results. If you don't recognize anyone, please say so; (Initial _____)
14. Whether or not you select someone, the police will continue to conduct an appropriate investigation into this matter. (Initial _____)

Initial estimates on the number of showups conducted in these jurisdictions suggested that more than a sufficient number of showup cases would be available to test our hypotheses. However, it became evident to us that many cases identified as having showups conducted were unusable due to various issues. For example, the case may have had missing written documentation about confidence or had no useful confidence statement,²² the case may have

²² For example, in some cases, the officer asked for a confidence statement and instead got a statement about the identification rather than one about confidence, such that the response to the question “how confident are you?” might have had a response such as “that was the hat he was wearing,” instead of something like “I am 99% sure” or “I am positive.” This problem may indicate that more training is needed for officers so that they are advised to re-state their request for a witness’ confidence level when the answer is non-responsive. It could also be due to policy; for example, the NRC report suggests confidence be collected in the witness’ own words, yet at least one study relied upon confidence categories that proved reliable (W.Wells et al., 2016).

been ineligible for review due to state restrictions (e.g., cases involving juveniles), or the case records may have been unavailable (e.g., state case expungement rules meant that some case files were no longer available²³), there may have been insufficient case information to be useful (e.g., the report may not have contained any information from the showup, even though there was an indication that a showup took place). While some may consider these factors to bias the results, they represent the legal and practical realities of conducting real-world research.

Unfortunately, latency, or the speed with which witnesses make their identification decision, was not a measure previously collected by either of our research sites. We also learned from some agencies during the course of time in which we were seeking LEA research partners, that it can be logistically difficult to identify a start time for a showup. For example, when an officer drives up to a scene, a witness may see the person matching their description from afar and begin to make verbal statements, even if instructed to wait until the officer asks a question. While it seems perfectly feasible to prevent this by advising a witness not to view the person until the officer tells them to do so, it may be practically difficult to avoid, without blocking the view of the witness or engaging in an alternative procedure. Nevertheless, since we had to rely on data that had already been collected, neither jurisdiction had complete records of latency, even when it was a recommended practice.

Overcoming the “Ground Truth” Problem

The inability to know the ground truth about whether the identified suspect is indeed the perpetrator of the crime remains a central concern for field-based eyewitness identification research, even when DNA evidence is available. However, the realities of actual cases are often

From a research perspective, translating a statement to a quantitative, categorical response is an imperfect science, so the present study relied on establishing categories, coding, and interrater agreement.

²³ States have different requirements about criminal case expungement and defendant requests for expungements.

far more nuanced than anything that can be constructed in a laboratory setting; yet this key limitation may weaken the argument for conducting research in the field.

To overcome this key limitation and test our hypotheses in the field, we needed to identify an approach that would allow us to estimate ground truth or serve as a close approximation. While some field studies of photo arrays allow for “known innocents” in the lineups, showups consist of only one individual, and it would not be practically or ethically feasible or efficient to provide a real-world witness with a known innocent as a showup suspect. Hence, it is generally not possible to validate whether a person identified by police as matching (or coming close to matching) the witness’ description of the perpetrator is the actual perpetrator.

While there are two other specific alternatives for capturing real-world ground truth – DNA and case outcomes (found guilty, pled guilty, not found guilty)–both have their own problems. DNA, for example, is only available in a fraction of the cases (primarily stranger sexual assaults and murders) and many jurisdictions still have large backlogs in DNA testing (see e.g., Wickenheiser, 2021). Moreover, while DNA evidence can unmistakably determine if a suspect was present at the scene, DNA alone may not prove that the person committed any crime much less the crime alleged (see e.g., Connors et al., 1996). Often, DNA evidence often fails to be conclusive. Nevertheless, the belief persists that DNA serves as “proof positive” of actual guilt or innocence. For instance, researchers have shown that jurors with a higher level of prosecution bias tend to overestimate the weight of weak DNA evidence (Smith & Bull, 2012).²⁴

The other alternative, using case outcomes may be reflective of ground truth, they can be influenced by other factors, such as plea bargaining and inadmissibility of evidence (despite its

²⁴ Still, there appears to be substantial value in DNA evidence for mitigating eyewitness inaccuracies. Finklea and Ebbesen (2007) found that when DNA was used as a validity check for ground truth, witnesses were wrong in just 5% of cases. Notably, in the 95% of cases in which witnesses were accurate in their identifications, two-thirds were for sexual assaults. The higher accuracy rate sexual assault cases may be explained by the prolonged exposure present such cases compared to briefer encounters, such as purse snatchings, robberies, or assaults.

probative value), legal standards, quality of prosecution, and witnesses' willingness and/or availability to testify. The length of criminal proceedings can also render case outcomes unfeasible for research.

As such, other means for establishing ground truth are necessary to advance science. Some of these approaches have been previously examined. Our approach in particular that has been shown promise involves examining and determining the amount, and more importantly, strength of the case evidence, independent of the eyewitness identification. With several potential types of evidence available in a given case, some evidence types and their quality are better than others in establishing the guilt or innocence of an identified suspect.

To help overcome the ground truth problem, we relied on an established instrument previously developed by one of the authors of this report.

Proxy for ground truth: Strength of Evidence Scale (SES). Amendola & Slipka (2009) oversaw the development of a comprehensive evidentiary-strength rating instrument named the Strength of Evidence Scale (SES). The SES is an objective and scientifically derived instrument used by key criminal justice decision-makers in this study to rate evidentiary strength in cases. This scale was developed for use in previous field research to validate the accuracy of eyewitnesses in field studies.

While prosecutors have more traditionally used ratings of evidentiary strength to prioritize cases for prosecution (see e.g., Jacoby et al., 1982), the use of evidentiary strength in estimating ground truth in eyewitness identification has been tested in studies (Behrman & Davey, 2001; Gould et al., 2012; Tollestrup et al., 1994). The SES was informed in part by a prominent effort at establishing evidentiary standards and a case evaluation system—the Bronx

County District Attorney case evaluation scheme for prioritizing cases for prosecution²⁵ (Jacoby et al., 1982). The primary limitation of the Bronx County District Attorney’s system was its reliance on a mathematical algorithm that was overly complex and left no room for prosecutor discretion and experience to play a role.

The SES, however, is primarily rooted in an approach to performance evaluation called Behaviorally Anchored Rating Scales (BARS), a method for making subjective evaluations more objective (Smith & Kendall, 1963). BARS are used across industries for performance evaluation as well as for scoring tests and promotional assessment centers, and for professional development (e.g., Debnath et al., 2015). The primary adaptation for our SES was to rate the strength and quality of the evidence in a criminal case, not the performance of individuals in work settings. In all other regards, the psychometric and structural properties remain similar.

This simplified SES tool, like BARS, allows for calibrating personal judgments based on agreed-upon criteria to establish exemplars of low and high scores, i.e., “anchors” at the outset. This differs from some forms of evaluation that rely entirely on subjective decisions that are based on various individual assumptions, personal biases, and beliefs, e.g., considering where a suspect lives as relevant to a decision about whether that person committed the crime. Through the use of a calibrated instrument demonstrating validity in research, these biases and beliefs are reduced, increasing precision as to what is “strong” or “weak” evidence, as well as eliminating non-relevant factors from consideration, e.g., the crime rate in that particular place, etc.

The inclusion and almost exclusive input²⁶ of police, prosecutors, defense attorneys, and judges (as key subject matter experts) in developing the instrument over 15 years ago, renders it

²⁵ The National Center for Prosecution Management and the National District Attorneys Association carried out the development of the case evaluation system at the request of Mario Merola, Bronx District Attorney (Jacoby et al., 1982).

²⁶ In the original development process, we also included scholars and others with backgrounds in criminology/criminal justice (especially policing), psychometrics and psychology.

particularly relevant to real-world cases and criminal justice proceedings. Because many extralegal factors influence the criminal justice process and affect case dispositions (Adams, 1983; Alderden & Ullman, 2012; Peterson et al., 1987), the instrument relies on evidentiary strength as a proxy for ground truth (actual guilt or innocence).

The scale covers six categories of evidence (physical, suspect statement, suspect history, victim characteristics, witness characteristics, and identification information) generated by the expert team, and comprises a 5-point scale for rating evidentiary strength ranging from 1 = evidence and/or information that is weak to 5 = evidence and/or information that is strong. The mid-point represents evidence that is neither weak nor strong. In **Figure 7**, we present an example of the rating scale for one of the rating dimensions (physical evidence).

Evidence for the predictive validity of the SES was initially generated from an experimental field study of the role of eyewitness identification information in evaluating evidentiary strength (Amendola et al., 2014, *The Austin Eyewitness Field Study*). In that study, the SES was used by actual criminal justice practitioners²⁷ to gauge the likelihood that suspects identified in eyewitness procedures were actually innocent or guilty based on the available evidence, independent of the identification outcome. Researchers found that higher evidentiary strength scores were associated with a greater likelihood of convictions on a 1 – 5 scale where 5 represented highly likely (mean of 4.34 for cases adjudicated guilty versus 2.5 for the non-prosecuted cases), indicating that these scores were probative of actual innocence, independent of judicial outcomes (Amendola & Wixted, 2015).

²⁷ These experts were mostly former prosecutors, defense attorneys, police investigators, and judges; some were still practicing in other jurisdictions. Teams of four (one from each expert group) independently rated actual previously-adjudicated cases where information about the identification decision and other necessary information was redacted and then participated in consensus discussions with their group to determine if they were influenced by others and wanted to change their prior ratings (this occurred infrequently, but primarily in cases where they had missed a detail in a case that was pointed out by another group member).

Figure 7

Physical Evidence Rating Scale, from *Strength of Evidence Scale* (Amendola & Slipka, 2009).

I. PHYSICAL EVIDENCE: Tangible items that directly link the suspect to the crime.

1 ← Evidence and/or information that is weak. | Evidence and/or information that is strong. → 5

FACTOR		FACTOR RATING
1. DNA - biological material recovered from crime scene	<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>DNA sample is contaminated (1.40)</p> <p>Fluid sample lacks sufficient quantity of DNA for testing (1.55)</p> </div> <div style="width: 45%;"> <p>Fluids from the crime scene prove joint presence of DNA for both victim AND suspect (4.46)</p> <p>Multiple DNA tests with consistent findings (4.69)</p> </div> </div>	<input style="width: 80px; height: 60px;" type="text"/>
2. Surveillance Tapes/ Photos from Crime Scene - images or photos that capture the crime	<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>Parts of surveillance tape are clear and parts are not (2.81)</p> <p>Videotape footage is within close proximity of the suspect committing the crime (3.69)</p> </div> <div style="width: 45%;"> <p>Tape provides a profile view of the suspect (3.64)</p> <p>Video captures the entire crime in real time (4.83)</p> </div> </div>	<input style="width: 80px; height: 60px;" type="text"/>
3. Fingerprints - fingerprint evidence recovered from crime scene	<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>Full fingerprint is smudged (1.70)</p> <p>Portion of fingerprint is smudged (2.51)</p> </div> <div style="width: 45%;"> <p>Suspect's fingerprint found at the crime scene (4.17)</p> <p>Suspect print found on weapon used in commission of crime (4.39)</p> </div> </div>	<input style="width: 80px; height: 60px;" type="text"/>
4. Wire Taps/ Audio Tapes - audio material that gives information about the crime	<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>No identifying information is provided on the tape (1.47)</p> <p>Audio tape has intermittent background noise (3.04)</p> </div> <div style="width: 45%;"> <p>Volume of audio tape is loud (3.42)</p> <p>Name of suspect is identified in audio taped conversation (3.59)</p> </div> </div>	<input style="width: 80px; height: 60px;" type="text"/>
5. Trace Evidence - including bite marks, tire marks, tools, etc	<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>Indistinguishable bite marks found on victim (1.73)</p> <p>Tire marks at crime scene are consistent with suspect's car (3.24)</p> </div> <div style="width: 45%;"> <p>Tool owned by suspect matches marks on victim's door (3.43)</p> <p>Casing from crime scene matches the casings from suspect's gun (4.24)</p> </div> </div>	<input style="width: 80px; height: 60px;" type="text"/>
6. Recovery of Items - such as drugs, guns, knives, and stolen items	<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>No shell casings found (1.90)</p> <p>Stolen item found in suspect's girlfriend's car (3.60)</p> </div> <div style="width: 45%;"> <p>Perp's high school class ring is recovered at crime scene and is from suspect's class, school, & year (3.48)</p> <p>Recovered bullet matches suspect's gun (4.40)</p> <p>Victim's driver's license is on the suspect (4.35)</p> </div> </div>	<input style="width: 80px; height: 60px;" type="text"/>
7. Other Miscellaneous Evidence	<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>Suspect's home and crime scene fall in the same cell tower radius. (1.95)</p> <p>Charges made on stolen card are made near the crime scene (3.55)</p> </div> <div style="width: 45%;"> <p>Suspect writing sample matches writing sample on a robbery note (3.77)</p> <p>Items are purchased online using the victim's card and are shipped to suspect's home (4.40)</p> </div> </div>	<input style="width: 80px; height: 60px;" type="text"/>
RATER # _____	PHYSICAL EVIDENCE OVERALL RATING	<input style="width: 80px; height: 60px;" type="text"/>

Additional research on the use of the SES for establishing ground truth (approximating guilt) and its validation has been demonstrated elsewhere (see e.g., Gould et al., 2012; Katzman & Kovera, 2022).

Study Protocols

For the purposes of this field study, the evidentiary strength scale was used to provide corroborating evidence, even if the evidence relied upon from the case file would not have been admissible in court, as it would add to the diagnosticity of evidence for the purposes of the research. To limit selection bias, we gathered records from successful showups in which the suspect was identified, as well as failed showups in which the suspect was not identified. This provided us with a range of decisions we could then compare to evidentiary strength ratings. Decisions ranged from not making an identification to making an identification with no confidence, bounded confidence, or high confidence.

Obtaining and coding confidence statements. As we learned from the Houston Police Department field study, a simple confidence scale works very well and is purportedly the best way to ask for confidence (W. Wells et al., 2016). In both sites, confidence statements had been collected in the victim/witness' own words. Subsequently, we coded the statements of the victim/witness into a 4-point scale as follows:

- 1 indicated “definitely not or not the person,”
- 2 indicated “possibly the person/somewhat confident,”
- 3 indicated “probably the person/confident to very confident,” and
- 4 indicated “definitely the person/highly confident (100% sure)”.

Protocols for rating evidentiary strength. As in the Austin Field Study (Amendola et al., 2014), we relied upon actual criminal justice officials²⁸ to rate the strength of the corroborating evidence using the SES as a proxy for ground truth (actual guilt or innocence). We conducted a four-hour training session for case evaluators. The training included a) an overview of the procedure for using the SES, b) a review of the categories of evidence (dimensions), c) a description of the types of evidence (factors) within each dimension, d) the rating scale anchors, e) the point values associated with each anchor (i.e., exemplar), and f) how to efficiently evaluate the evidence. In the final segment of training, we provided the evaluators with practice cases to independently evaluate. After their ratings were complete, we facilitated a discussion of the cases for the evaluators to establish consensus, if possible, by calibrating their approaches to evaluating the evidence. Of course, agreement was not required, just discussion of the cases.

In both jurisdictions, we provided the case evaluators with case files containing police reports stripped of personally identifiable information (where required), case dispositions and showup identifications so as not to influence their determination of the strength of the corroborating evidence against the specific suspect. We then asked the raters to use the SES to evaluate the strength of the evidence independently, across the same cases. After receiving the independent ratings from each evaluator, we identified score discrepancies of two or more points, but only for those cases in which one evaluator considered the evidence to be weak while the other considered it to be strong (i.e., a 1 and 4, a 2 and 4, a 1 and 5, or a 2 and 5, as 3 was neutral) and vice versa.²⁹ As in the training, we then facilitated a discussion among the two evaluators to discuss any discrepant ratings. For the few cases in which the ratings diverged by

²⁸ In this case, a former prosecutor and police detective.

²⁹ It could be argued that every case should have been discussed, even when there was agreement, but such a process would require additional time and associated expense, while not likely leading to a meaningful change. As such, the use of an average rating between two close scores, e.g. a 2 (relatively weak) and a 3 (neither weak nor strong), or a 1 (weak evidence) and 3 would not significantly change the rating from weak to strong or vice versa.

two or more points as described above, it was determined that either one of the evaluators misinterpreted the case details or had a different understanding of the details provided. We then allowed the evaluators to revise their initial rating or keep it for our final scoring. In some cases, one or both evaluators reduced or increased their rating by one point, at their own will. We computed the mean across the pair of case evaluators for the final analysis. Finally, we analyzed the ratings in conjunction with the coded victim and/or witnesses' statements of confidence.

Analysis

Statistical Power

Our power analysis was conducted prior to our knowledge that we could not test for latency in our field study. Originally, we determined that approximately 350 showups would be more than sufficient to detect meaningful differences. We eliminated showups for which confidence was not recorded in the witness or victim's own words, resulting in 80 showups in Hartford (for the period from 2014 through 2021) and 73 showups in Denver (for the period from 2021 through 2023) or 153 total usable showups. Nevertheless, since we were not able to test our second hypothesis about latency, analytical power increased two-fold. Thus, when considering only a single predictor, we would attain the same level of power with just 175 cases. Although we only had 153 (pooled) usable cases, we conducted the analysis as planned, with the acknowledgment that if the findings were significant, the obtained effect sizes may be slightly overestimated.

Analysis Plan

We developed logistic and multinomial regression models and applied them to the data, which consisted of 153 showups and their associated cases. We also included hierarchical models since two agencies participated. Although we examined data from all showups, our

primary focus was on showups that involved a witness and offender who were strangers and for which agency protocols were followed.

We also examined the factors that predicted the witness' decision to choose. The decision to select (or not select) the individual in the showup was accompanied by the level of confidence reported by the witness. We examined the level of confidence reported by the witness using confidence-accuracy characteristic analysis (CAC; Mickes, 2015; Seale-Carlisle et al., 2019). As mentioned in Study One, CAC analysis is a measure of reliability because it allows researchers to estimate the likelihood that the suspect identified is guilty. Evidentiary strength ratings for each case allowed us to subdivide suspect choices as a function of the degree of corroborating evidence from the expert ratings from the SES (Amendola & Slipka, 2009).

Results

We applied several analyses to the combined showup data from Denver and Hartford. As described in our methods section, the open-ended confidence statements provided by the eyewitnesses were coded on a scale from 1 to 4 where:

1 = Definitely not the person,

2 = Possibly the person,

3 = Probably the person, and

4 = Definitely the person

We then conducted an analysis of evidentiary strength based on all non-eyewitness evidence. As stated earlier, the case evaluators were blind to the eyewitness evidence. This analysis resulted in an average rating of 1–5 for each case, where 1 = low evidentiary strength of suspect guilt and 5 = high evidentiary strength of guilt.

Our goal was to determine whether there was a relationship between the confidence

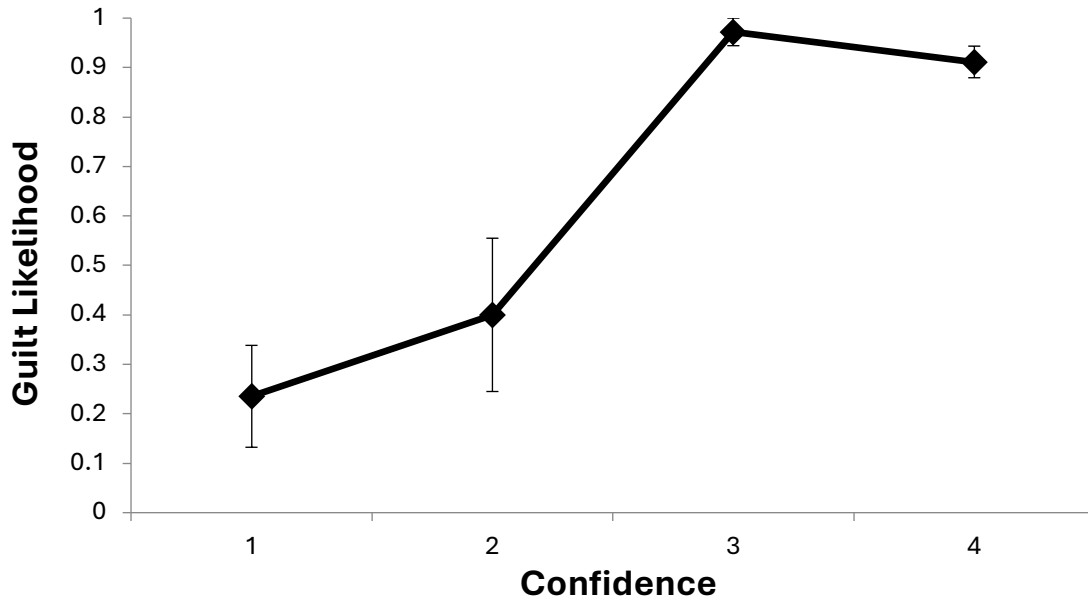
rating and evidentiary strength. In other words, we wanted to determine whether suspects identified with higher confidence tended to have stronger non-eyewitness evidence (other evidence) of guilt. This method was our way of estimating potential target-presence (i.e., higher evidentiary strength of guilt ratings) versus target-absence (i.e., lower evidentiary strength of guilt ratings) in the showups so that we could calculate the likelihood of guilt as the basis for our confidence-accuracy (i.e., confidence-evidentiary strength) analyses.

Our first analysis was a simple Pearson correlation between confidence levels that ranged from 1–4 and evidentiary strength ratings that ranged from 1–5. There was a correlation between these two measures, $r(152) = .56$, $R^2 = .316$, an effect size considered strong (Cohen, 1988). To place the strength of this relationship in context, in their review of the confidence-accuracy relationship in eyewitness identification studies, Wixted and G. Wells (2017) argued that the point-biserial correlation of .41 from the Sporer et al. (1995) meta-analysis could be considered to indicate a “very strong confidence-accuracy relationship” (p. 22). We also applied a linear regression model, which supported the strong relationship that we found, $F(1, 151) = 69.61$, $p < .001$.

Next, we created CAC curves (Mickes, 2015) to depict the overall relationship between confidence and likelihood of guilt (High Guilt Ratings/[High Guilt Ratings + Low Guilt Ratings]). “High Guilt Ratings” were defined as cases yielding an evidentiary strength rating of 4–5, and “Low Guilt Ratings” were defined as cases yielding an evidentiary strength of 1–2. For this analysis, we excluded all cases with a rating of 3 or 3.5 because of the greater ambiguity regarding potential guilt, which accounted for only 11 out of the 153 total cases. A more extreme analysis focusing only on cases with evidentiary strength ratings of 5 versus 1–1.5 was not possible due to too little data (only 25 cases). See **Figure 8** for the CAC curve.

Figure 8

Confidence-accuracy characteristic curve (with standard error bars)



There are two aspects of CAC space on which to focus. First is the overall relationship between confidence and likelihood of guilt, which was positively sloped as expected, as shown in **Figure 8**, with a lower guilt likelihood for low-confidence decisions and a higher likelihood for high-confidence decisions based on the showups from Denver and Hartford. Second, is guilt likelihood associated with the highest confidence. These are important IDs in real-world cases, and ideally, accuracy would be very high (i.e., above 90%) to support the high-confidence identifications from the eyewitnesses. This pattern was also supported in our data. Yet there was one anomaly in terms of the difference in accuracy between a moderate and high confidence level not reflected in our lab study or others, as described below.

One issue to note is that, in our rating system, a confidence rating of 1 (“definitely not the

person”) is a rejection and, as such, does not belong in CAC space with IDs. We chose to portray the full confidence scale in **Figure 8** for transparency, but a more appropriate analysis of the space (to focus on the IDs) would involve only confidence ratings of 2–4. Fortunately, the conclusion is still the same: suspects identified with low confidence were not likely guilty, but those identified with medium or high confidence were likely to be guilty.

In addition to relying upon the visual inspection of the CAC curve, we also tested these relationships statistically by applying two-tailed z -tests to compare different pairs of points along the curve. There was no difference in guilt likelihood between confidence rating 1 (i.e., rejections) versus 2 (low confidence IDs), $z = 0.91$, $p = 0.36$. There was a large increase in guilt likelihood from confidence rating 2 (low) to 3 (medium), $z = 4.42$, $p < 0.001$, h (effect size) = 1.42. Lastly, the anomaly we referenced above was that there was no difference between confidence rating 3 (medium) and 4 (high), $z = 1.16$, $p = 0.25$, suggesting that a designation as “*definitely the person*: may not increase likelihood of guilt beyond a “*probably the person*” (moderately high confidence) classification.

Our final approach to analyzing these data was to apply calibration analysis (Juslin et al., 1996) to reveal a resolution statistic (also known as discrimination). A resolution statistic indicates the ability of confidence to discriminate between correct (suspect likely guilty) and incorrect (suspect likely innocent) ID decisions. Specifically, we calculated the Adjusted Normalized Resolution Index or ANRI (Yaniv et al., 1991; also known as Adjusted Normalized Discrimination Index, see Carlson et al., 2016; Vredeveldt & Sauer, 2015; Whittington et al., 2019), which ranges from 0–1 with higher values indicating greater resolution/discrimination.

We calculated ANRI for both the full range of data depicted in **Figure 8** above (.41) and for confidence levels of 2–4 to focus solely on IDs (.22). The greater resolution when including

the full confidence range is simply due to the greater range in guilt likelihood: $0.91 - 0.235 = 0.675$, whereas focusing on the IDs only (confidence of 2–4) reduces the range to .51. The resolution of 0.22 for IDs is commensurate with the eyewitness ID literature using fair simultaneous lineups (e.g., 0.23 from Carlson et al., 2016; 0.22 from Palu et al., 2024).

Discussion

Limitations

The primary limitation of the field studies was the lack of fully available data from cases in which showups took place. Given that this is generally out of the control of researchers, agencies should be encouraged to ensure officers document all aspects of the case. It is possible that some of the details that were missing in some cases were provided in body-worn camera (BWC) footage that was not available to researchers in this study. In future studies, researchers should attempt to get recordings associated with showups, by identifying jurisdictions without restrictions that would prevent the ability to obtain BWC footage, as many agencies have such restrictions regarding their use and release.

Relatedly, the other key limitation of the study is the lower than anticipated number of cases that reduced overall power, despite only testing one of the two original hypotheses. This can sometimes lead to an overestimate of the effect sizes. Given the relatively strong effect sizes obtained in the field study, it is unlikely that the results of reduced power would reduce the effect sizes substantially, however.

A final limitation of the study was the collection of previously collected data from the field. In the future, if time permits, it would be ideal to conduct a replication in a jurisdiction where a standardized latency measure has been employed. In the event that identifying such an agency is not possible, then it would be useful for researchers to conduct a study in which the

agency begins collecting latency procedures along with confidence statements in showups and then follow these cases through case disposition, provided there is sufficient time.

Conclusions from Study Two

We investigated the confidence-accuracy relationship in field-showup identifications using a proxy for ground truth: the Strength of Evidence Scale. By relying on a few analyses, a simple correlation, CAC analysis and associated significance testing, and an index of discrimination (resolution statistic). Using this approach for sorting correct from incorrect decisions in the real-world field study data provides important converging evidence vital to our evaluation of showup decisions. Our findings reveal that confidence is indicative of accuracy in field showups in two jurisdictions in two different regions in the U.S. This is important for both confident identifications and confident rejections. Importantly and consistent with prior lab data, showup rejections also provide evidence of innocence that should not be ignored. In many exoneration cases, the witnesses often rejected the lineup or showup during the first viewing, but subsequent tests were then mistakenly used to convict innocent parties.

If this pattern continues to hold across a range of field-based experiments, identifications stemming from showup procedures may be considered reliable and accurate, provided administration procedures designed to reduce suggestibility or other biases are employed.

The fact that there was no significant difference in accuracy for moderately high to very high confidence is notable, especially because lab study research has demonstrated such a difference. There is a possibility that defense attorneys can establish a “reasonable doubt” with anything other than a 100% or “absolutely certain” identification despite no additional incremental value in an absolutely certain confidence rating over moderately high confidence (“probably the person”) in terms of its ability to predict actual guilt.

Moderate to high confidence identifications in showups both appear to be probative of guilt and thus confident witness identifications may improve the reliability of showup identifications as well as to establish the utility of showups as, at a minimum, a reasonable tool in the early stages of police investigations of crimes. In agencies that obtain confidence statements, it is unlikely that cases with no or low confidence would be pursued without other useful evidence. After all, basic estimator variables (like viewing conditions, duration, stress, weapon presence, etc.) often prevent witnesses from being able to make strong positive identifications even if the other evidence in the case is probative. Hence, the implications for collecting confidence and obtaining other than at least moderate confidence would not be of consequence to an investigation anyhow.

Needless to say, we do not recommend the use of showups as a solitary investigative tool. Instead, we see showups as a first step in an investigation when little other evidence is available or as an additional probative factor after other evidence has been gathered. The efficacy of collecting initial confidence statements in both photo arrays and showups (and latency in at least photo arrays) has been demonstrated in both the lab and field studies, consistent with current research and current recommendations by many scholars and other groups that have established practice guidelines and recommendations. It is expected that confidence (and latency where it can be established) will strengthen cases against guilty suspects while reducing the likelihood of wrongful convictions of innocent persons.

Overall Conclusions

In sum, the lab study's findings indicate that both confidence and response time are predictive of accuracy in live showups. The field study's findings similarly indicate that confidence reflects accuracy in live showups. Latency was not captured in the field studies

because of the inability to identify a specific start time when witnesses were brought (typically by vehicle) to the suspect in the real-world setting. As such, the latency-accuracy relationship could not be tested in the field.

If these findings are replicated in additional lab and field studies examining showups, the evidence arising from this project will help establish the probative nature of confidence statements, aid in the development of improved law enforcement policies, and help strengthen cases against guilty suspects while reducing the likelihood of wrongful prosecutions and convictions of innocent people.

References

- Adams, K. (1983). The effect of evidentiary factors on charge reduction. *Journal of Criminal Justice*, 11, 525-537. [https://doi.org/10.1016/0047-2352\(83\)90005-3](https://doi.org/10.1016/0047-2352(83)90005-3)
- Alderden, M. A., & Ullman, S. E. (2012). Creating a more complete and current picture: Examining police and prosecutor decision-making when processing sexual assault cases. *Violence Against Women*, 18(5), 525-551. <https://doi.org/10.1177/1077801212453867>
- Amendola, K. L., Mickes, L., Olson, M. V., Gao, J., Gronlund, S., Yang, Y. Carlson, C., and Chen, X. (2025). Confidence, latency, and accuracy in eyewitness identifications made from showups: Nationwide survey of eyewitness practices. Final report to the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. *National Policing Institute, Arlington, VA.*
- Amendola, K. L., & Slipka, M. G. (2009). Strength of evidence scale. *Unpublished instrument. Police Foundation, Washington, DC.* <https://www.policinginstitute.org/wp-content/uploads/2020/04/About-the-Strength-of-Evidence-Scale.pdf>
- Amendola, K. L., Valdovinos, M. D., Hamilton, E. E., Slipka, M. G., Sigler, M., & Kaufman, A. (2014). Photo arrays in eyewitness identification procedures: Presentation methods, influence of ID decisions on experts' evaluations of evidentiary strength, and follow-up on the AJS Eyewitness ID Field Study. Washington, DC: Police Foundation. *Washington, DC: Police Foundation. Report under Revision.* <https://www.publicsafety.gc.ca/lbrr/archives/cnmcs-plcng/cn31236-eng.pdf>
- Amendola, K. L., & Wixted, J. T. (2015). Comparing the diagnostic accuracy of suspect identifications made by actual eyewitnesses from simultaneous and sequential lineups in

a randomized field trial. *Journal of Experimental Criminology*, 11, 263-284.

<https://doi.org/10.1007/s11292-014-9219-2>

Behrman, B. W., & Davey, S. L. (2001). Eyewitness identification in actual criminal cases: An archival analysis. *Law and Human Behavior*, 25(5), 475-491.

<https://doi.org/10.1023/A:1012840831846>

Berkowitz, S. R., Garrett, B. L., Fenn, K. M., & Loftus, E. F. (2020). Convicting with confidence? Why we should not over-rely on eyewitness confidence. *Memory*, 30(1), 10-15. <https://doi.org/10.1080/09658211.2020.1849308>

Bertelsman, S. P. (2022). Defending due process: The case for abolishing the show-up line-up. *Washington University Journal of Law and Policy*, 68, 245.

<https://heinonline.org/HOL/LandingPage?handle=hein.journals/wajlp68&div=13&id=&page=>

BJS, see: 1. Langan & Dawson (1993), 2. Durose & Langan (2004), 3. Rosenmerkel et al. (2009).

Blomberg, T. G. (2021). Translational Criminology and Politics Presentation. Justice Research and Statistics Association's (JRSA) Virtual Training. Florida State University.

<https://criminology.fsu.edu/sites/g/files/upcbnu3076/files/JRSA%20Blomberg%202021-11-30-21.pdf>

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

Brooks, N. (1983). *Police guidelines: Pretrial eyewitness identification procedures*. Ottawa: Law Reform Commission of Canada.

https://publications.gc.ca/collections/collection_2022/jus/J32-6-5-1983-eng.pdf

Carlson, C. A., Hemby, J. A., Wooten, A. R. *et al.* (2021). Testing encoding specificity and the diagnostic feature-detection theory of eyewitness identification, with implications for show-ups, lineups, and partially disguised perpetrators. *Cognitive Research: Principles & Implications*, 6, 14. <https://doi.org/10.1186/s41235-021-00276-3>

Carlson, C. A., Young, D. F., Weatherford, D. R., Carlson, M. A., Bednarz, J. E., & Jones, A. R. (2016). The influence of perpetrator exposure time and weapon presence/timing on eyewitness confidence and accuracy. *Applied Cognitive Psychology*, 30(6), 898-910. <https://doi.org/10.1002/acp.3275>

Cassell, P. G. (2018). Overstating America's wrongful conviction rate: Reassessing the conventional wisdom about the prevalence of wrongful convictions. *Ariz. L. Rev.*, 60, 815. <https://dc.law.utah.edu/cgi/viewcontent.cgi?article=1130&context=scholarship>

Cicchini, M. D., & Easton, J. G. (2010). Reforming the law on show-up identifications. *The Journal of Criminal Law and Criminology*, 381-413. <https://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=7352&context=jclc>

Clark, S. E., Moreland, M. B., & Larson, R. P. (2018). Legitimacy, procedural justice, accuracy, and eyewitness identification. *UC Irvine L. Rev.*, 8, 41. <https://heinonline.org/HOL/LandingPage?handle=hein.journals/ucirvlre8&div=7&id=&page=>

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers. <https://www.utstat.toronto.edu/brunner/oldclass/378f16/readings/CohenPower.pdf>

- Colloff, M. F., & Wixted, J. T. (2020). Why are lineups better than showups? A test of the filler siphoning and enhanced discriminability accounts. *Journal of Experimental Psychology: Applied*, 26(1), 124. <https://doi.org/10.1037/xap0000218>
- Connors, E., Lundregan, T., Miller, N., & McEwen, T. (1996). *Convicted by juries, exonerated by science: Case studies in the use of DNA evidence to establish innocence after trial*. Research Report. NCJ 161258. Washington DC: National Institute of Justice. Office of Justice Programs, U.S. Department of Justice. <https://www.ojp.gov/pdffiles/dnaevid.pdf>
- Debnath, S. C., Lee, B. B., & Tandon, S. (2015). Fifty years and going strong: What makes Behaviorally Anchored Rating Scales so perennial as an appraisal method?. *International Journal of Business and Social Science*, 6(2).
<https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=9c381f049faa17915d904472fa09f67ed6b2ec46>
- 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(5), 951. https://www.researchgate.net/profile/David-Dunning-2/publication/11067061_Automaticity_and_Eyewitness_Accuracy_A_10-to_12-Second_Rule_for_Distinguishing_Accurate_from_Inaccurate_Positive_Identifications/links/55ef043108aef559dc44aec2/Automaticity-and-Eyewitness-Accuracy-A-10-to-12-Second-Rule-for-Distinguishing-Accurate-From-Inaccurate-Positive-Identifications.pdf
- Durose, M. R. & Langan, P. A. (2004). "Felony sentences in state courts, 2002." *Bureau of Justice Statistics Bulletin*. NCJ 206916. Washington, DC Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice.
<https://bjs.ojp.gov/content/pub/pdf/fssc02.pdf>

- Dysart, J. E., & Lindsay, R. C. L. (2007). The effects of delay on eyewitness identification accuracy: Should we be concerned?. In *The Handbook of Eyewitness Psychology: Volume II* (pp. 375-390). Psychology Press. https://www.routledge.com/The-Handbook-of-Eyewitness-Psychology-Volume-II-Memory-for-People/Lindsay-Ross-Read-Toglia/p/book/9780367463052?srsId=AfmBOoq-yt9rW2dwXGa7wg5ytNiMMZ_1mQPXuA2NnHKzzJEh_GAcsUTE
- Eisen, M. L., 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), 1. <https://doi.org/10.1037/law0000115>
- Eisen, M. L., Ying, R. C., Chui, C., & Swaby, M. A. (2022). Comparing witness performance in the field versus the lab: How real-world conditions affect eyewitness decision-making. *Law and Human Behavior*, 46(3), 175-188. <https://doi.org/10.1037/lhb0000485>
- Finklea, K. M., & Ebbesen, E. B., (2007). *Eyewitness accuracy in the real world: DNA evidence as “ground truth” for eyewitness accuracy rates*. Unpublished manuscript.
- Friend, K. B., Gordon, M., Scarbrough, B., Collins, D., Fritz, K., Smoot, S., ... & Joyce, N. (2020). Sentinel event reviews in the criminal justice system: a review of the literature. *Criminal Justice Studies*, 33(4), 337-353. <https://doi.org/10.1080/1478601X.2020.1741227>
- Gonzalez, R., Ellsworth, P. C., & Pembroke, M. (1993). Response biases in lineups and showups. *Journal of Personality and Social Psychology*, 64(4), 525. <https://psycnet.apa.org/buy/1993-27878-001>
- Green, L. W., Ottoson, J. M., Garcia, C., & Hiatt, R. A. (2009). Diffusion theory and knowledge dissemination, utilization, and integration in public health. *Annual Review of Public*

- Health*, 30(1), 151-174. <https://doi.org/10.1146/annurev.publhealth.031308.100049>
- Gronlund, S. D., & Benjamin, A. S. (2018). The new science of eyewitness memory. In *Psychology of Learning and Motivation* (Vol. 69, pp. 241-284). Academic Press. <https://doi.org/10.1016/bs.plm.2018.09.006>
- Gronlund, S. D., Carlson, C. A., Neuschatz, J. S., Goodsell, C. A., Wetmore, S. A., Wooten, A., & Graham, M. (2012). Show-ups versus lineups: An evaluation using ROC analysis. *Journal of Applied Research in Memory and Cognition*, 1, 221-228. <https://doi.org/10.1016/j.jarmac.2012.09.003>
- Gould, J. B., Carrano, J., Leo, R., & Young, J. (2012). *Predicting Erroneous Convictions: A Social Science Approach to Miscarriages of Justice*. (Report for Award No. 2009-IJ-CX-4110). Washington, DC: National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. <https://www.ojp.gov/pdffiles1/nij/grants/241389.pdf>
- Gross, S. R. & Shaffer, M. (2012). Exonerations in the U.S., 1989 – 2012. National Registry of Exonerations. https://www.law.umich.edu/special/exoneration/documents/exonerations_us_1989_2012_full_report.pdf
- Holdstock, J. S., Dalton, P., May, K., Boogert, S. & Mickes, L. (2022). Lineup identification in young and older witnesses: Does describing the criminal help or hinder?" *Cognitive Research: Principles and Implications*, 7(51). <https://doi.org/10.1186/s41235-022-00399-1>
- Hollway, J. F., & Grunwald, B. (2019). Applying sentinel event reviews to policing. *Criminology & Public Policy*, 18(3), 705-730. <https://doi.org/10.1111/1745-9133.12457>

International Association of Chiefs of Police (2016). Eyewitness Identification: Model Policy, Concepts & Issues Paper, and Need to Know. <https://www.theiacp.org/resources/policy-center-resource/eyewitness-identification>

Jacoby, J. E., Mellon, L. R., Ratledge, E. C., and Turner, S. T. (1982). *Prosecutorial Decision-making: A National Study*. (Report for Award No. 79-NI-AX-0006 and 79-NI-AX-0034). NCJ 79227. Washington DC: National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. <https://www.ojp.gov/pdffiles1/Digitization/79227NCJRS.pdf>

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(5), 1304-1316. <https://psycnet.apa.org/buy/1996-05780-017>

Katzman, J. A., & Kovera, M. B. (2022). Evidence strength (insufficiently) affects police officers' decisions to place a suspect in a lineup. *Law and Human Behavior*, 46(1), 30–44. <https://doi.org/10.1037/lhb0000476>

Katzman, J. A., & Kovera, M. B. (2024). 8 Police Decisions Involved in Collecting Eyewitness Identification Evidence. *The Cambridge Handbook of Psychology and Legal Decision-Making*, 117. https://www.google.com/books/edition/The_Cambridge_Handbook_of_Psychology_and_Akn8EAAAQBAJ?hl=en&gbpv=1&dq=kovera+role+of+eyewitness+identification&pg=PA117&printsec=frontcover

Key, K. N., Cash, D. K., Neuschatz, J. S., Price, J., Wetmore, S. A., & Gronlund, S. D. (2015). Age differences (or lack thereof) in discriminability for lineups and

showups. *Psychology, Crime & Law*, 21(9), 871-889.

<https://doi.org/10.1080/1068316X.2015.1054387>

Klobuchar, A., Steblay, N. K. M., & Caligiuri, H. L. (2006). Improving eyewitness identifications: Hennepin County's blind sequential lineup pilot project. *Cardozo Pub. L. Pol'y & Ethics J.*, 4, 381.

<https://heinonline.org/HOL/LandingPage?handle=hein.journals/cardplp4&div=28&id=&page=>

Kovera, M. B. (2024). The role of suspect development practices in eyewitness identification accuracy and racial disparities in wrongful conviction. *Social Issues and Policy Review*, 18(1), 125-147. <https://doi.org/10.1111/sipr.12102>

Langan, P. A. & Dawson, J. M. (1993). Felony sentences in state courts, 1990. *Bureau of Justice Statistics Bulletin*, NCJ 140186. Washington, DC: Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice.

<https://bjs.ojp.gov/library/publications/felony-sentences-state-courts-1990>

Loeffler, C. E., Hyatt, J. & Ridgeway, G. (2019). Measuring self-reported wrongful convictions among prisoners. *Journal of Quantitative Criminology*, 35, 259-286 (2019).

<https://doi.org/10.1007/s10940-018-9381-1>

McGarrell, E. F., Hipple, N. K., & O'Brien, M. (2023). Sentinel event reviews: applications in criminal justice settings. *Journal of Crime and Justice*, 46(5), 563-578.

<https://doi.org/10.1080/0735648X.2023.2184412>

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(2),

93-102. <https://doi.org/10.1016/j.jarmac.2015.01.003>

Mickes, L., Clark, S. E., & Gronlund, S. D. (2017). Distilling the confidence-accuracy message:

A comment on Wixted and Wells (2017). *Psychological Science in the Public*

Interest, 18(1), 6-9. <https://doi.org/10.1177/1529100617699240>

Mickes, L., Hwe, V., Wais, P. E., & Wixted, J. T. (2011). Strong memories are hard to scale.

Journal of Experimental Psychology: General, 140(2), 239-257.

<https://doi.org/10.1037/a0023007>

Mickes, L., Seale-Carlisle, T. M., Chen, X., & Boogert, S. (2023). pyWitness 1.0: A Python

eyewitness identification analysis toolkit. *Behavior Research Methods*.

<https://doi.org/10.3758/s13428-023-02108-2>

Milde, J. P. (2019). Bare Necessity: Simplifying the Standard for Admitting Showup

Identifications. *Boston College of Law Review*, 60, 1771.

<https://core.ac.uk/download/pdf/222446086.pdf>

Morris, Z. S., Wooding, S., & Grant, J. (2011). The answer is 17 years, what is the question:

understanding time lags in translational research. *Journal of the Royal Society of*

Medicine, 104(12), 510-520.

<https://journals.sagepub.com/doi/pdf/10.1258/jrsm.2011.110180>

National Institute of Justice (US). Technical Working Group for Eyewitness Evidence.

(1999). *Eyewitness Evidence: A Guide for Law Enforcement*. NCJ 178240. Washington,

DC: National Institute of Justice, Office of Justice Programs, U.S. Department of

Justice,. <https://nij.ojp.gov/library/publications/eyewitness-evidence-guide-law->

[enforcement](https://nij.ojp.gov/library/publications/eyewitness-evidence-guide-law-enforcement)

National Research Council (2014). *Identifying the Culprit: Assessing Eyewitness Identification*.

Washington, DC: The National Academies Press.

<https://nap.nationalacademies.org/catalog/18891/identifying-the-culprit-assessing-eyewitness-identification>

Neuschatz, J. S., Wetmore, S. A., Key, K. N., Cash, D. K., Gronlund, S. D., & Goodsell, C. A. (2016). A comprehensive evaluation of showups. *Advances in Psychology and Law: Volume 1*, 43-69. https://doi.org/10.1007/978-3-319-29406-3_2

NIJ (1999), see National Institute of Justice

Norman, D. A. (2010). The Research-Practice Gap: The Need for Translational Developers. *Interactions*, 17(4), 9-12. <https://doi.org/10.1145/1806491.1806494>

Palu, A., Raidvee, A., Murnikov, V., & Kask, K. (2024). The effect of surgical masks on identification decisions from masked and unmasked lineups. *Psychiatry, Psychology and Law*, 31(5), 932-962. <https://doi.org/10.1080/13218719.2023.2242435>

Peterson, J. L., Ryan, J. P., Houlden, P. J., & Mihajlovic, S. (1987). The uses and effects of forensic science in the adjudication of felony cases. *Journal of Forensic Sciences*, 32(6), 1730-1753. <https://doi.org/10.1520/JFS11231J>

Police Executive Research Forum (PERF). (2013). A national survey of eyewitness identification processes in law enforcement agencies. (Report for Award No 2010-IJ-CX0032).

Washington, DC: National Institute of Justice, Office of Justice Programs, U.S. Department of Justice.

https://www.policeforum.org/assets/docs/Free_Online_Documents/Eyewitness_Identification/a%20national%20survey%20of%20eyewitness%20identification%20procedures%20in%20law%20enforcement%20agencies%202013.pdf

Report of the Special Master, State v. Henderson, No. A-8– 08. (2008, September).

<https://www.physics.smu.edu/pseudo/Eyewitness/NJreport.pdf>

Rogers, E. M. (1962). *Diffusion of Innovations*. New York: Free Press of

Glencoe. <https://teddykw2.wordpress.com/wp-content/uploads/2012/07/everett-m-rogers-diffusion-of-innovations.pdf>

Rosenmerkel, S., Durose, M. & Farole, D. (2009). *Felony sentences in state courts, 2006*.

Statistical Tables. NCJ 226846. Washington, DC: Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice.

<https://bjs.ojp.gov/content/pub/pdf/fssc06st.pdf>

Sauerland, M., Sagana, A., & Sporer, S. L. (2012). Assessing nonchoosers' eyewitness identification accuracy from photographic showups by using confidence and response times. *Law and Human Behavior*, 36(5), 394. <https://doi.org/10.1037/h0093926>

Sauerland, M., Sagana, A., Sporer, S. L., & Wixted, J. T. (2018). Decision time and confidence predict choosers' identification performance in photographic show-ups. *PLoS ONE* 13(1): e0190416. <https://doi.org/10.1371/journal.pone.0190416>

Sauerland, M., & Sporer, S. L. (2009). Fast and confident: postdicting eyewitness identification accuracy in a field study. *Journal of Experimental Psychology: Applied*, 15(1), 46-62. <https://doi.org/10.1037/a0014560>

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(4), 420-428. <https://doi.org/10.1016/j.jarmac.2019.09.003>

Sjöberg, M. P. (2016). The show-up identification procedure: A literature review. *Open Journal of Social Sciences*, 4(1), 86-95. <https://doi.org/10.4236/jss.2016.41012>

- Smith, L. L., & Bull, R. (2012). Identifying and measuring juror pre-trial bias for forensic evidence: Development and validation of the Forensic Evidence Evaluation Bias Scale. *Psychology, Crime & Law*, 18(9), 797-815. <https://doi.org/10.1080/1068316X.2011.561800>
- Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology*, 47(2), 149-155. <https://doi.org/10.1037/h0047060>
- 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(3), 315-327. <https://doi.org/10.1037/0033-2909.118.3.315>
- Stoddard, G. D. (1935). The Lag between Practice and Research. *The Phi Delta Kappan*, 17(6), 180–182. <http://www.jstor.org/stable/20258410>
- Stovall v. Denno, 388 U.S. 293, 302 (1967). <https://supreme.justia.com/cases/federal/us/388/293/>
- Suggs, D. L. (1979). The use of psychological research by the judiciary: Do the courts adequately assess the validity of the research?. *Law and Human Behavior*, 3(1-2), 135.
- Taylor, S. E., Crocker, J., Fiske, S. T., Sprinzen, M., & Winkler, J. D. (1979). The generalizability of salience effects. *Journal of Personality and Social Psychology*, 37(3), 357-368. <https://doi.org/10.1037/0022-3514.37.3.357>
- Tekin, E., DeSoto, K. A., Wixted, J. H., & Roediger, H. L. (2021). Applying confidence accuracy characteristic plots to old/new recognition memory experiments. *Memory*, 29, 427-443. <https://doi.org/10.1080/09658211.2021.1901937>
- Tollestrup, P. A., Turtle, J. W., & Yuille, J. C. (1994). Actual victims and witnesses to robbery and fraud: An archival analysis. *Adult Eyewitness Testimony: Current Trends and Developments*, 144-160. <https://www.ojp.gov/ncjrs/virtual-library/abstracts/actual->

[victims-and-witnesses-robbery-and-fraud-archival-analysis](#)

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80(5), 352-373. <https://doi.org/10.1037/h0020071>

United States v. Newman, 144 F.3d 531, 535 (7th Cir. 1998). <https://case-law.vlex.com/vid/u-s-v-newman-893142736>

Valentine, T., Davis, J. P., Memon, A., & Roberts, A. (2011). Live show-ups and their influence on a subsequent video lineup. *Applied Cognitive Psychology*, 26, 1-23.

<https://doi.org/10.1002/acp.1796>

Vredeveltdt, A., & Sauer, J. D. (2015). Effects of eye-closure on confidence-accuracy relations in eyewitness testimony. *Journal of Applied Research in Memory and Cognition*, 4(1), 51-58. <https://doi.org/10.1016/j.jarmac.2014.12.006>

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(1), 3-36.

<https://doi.org/10.1037/lhb0000359>

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads.

Law and Human Behavior, 22(6), 603. <https://doi.org/10.1023/A:1025750605807>

Wells, W., Campbell, B., Li, Y., & Swindle, S. (2016). The characteristics and results of eyewitness identification procedures conducted during robbery investigations in Houston, TX. *Policing: An International Journal*, 39(4), 601-619. <https://doi.org/10.1108/PIJPSM-10-2015-0124>

[10-2015-0124](#)

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(1), 8-14.
<https://doi.org/10.1016/j.jarmac.2014.07.003>
- Whittington, J. E., Carlson, C. A., Carlson, M. A., Weatherford, D. R., Krueger, L. E., & Jones, A. R. (2019). Asking an eyewitness to predict their later lineup performance could harm the confidence–accuracy relationship. *Applied Cognitive Psychology*, 34(1), 119-131.
<https://doi.org/10.1002/acp.3599>
- Wickenheiser, R. A. (2021). The value of forensic DNA leads in preventing crime and eliminating the innocent. *Forensic Science International: Synergy*, 3, 100201.
<https://doi.org/10.1016/j.fsisyn.2021.100201>
- Wilson, B. M., Seale-Carlisle, T. M., & Mickes, L. (2018). The effects of verbal descriptions on performance in lineups and showups. *Journal of Experimental Psychology: General*, 147(1), 113-124. <https://doi.org/10.1037/xge0000354>
- Wixted, J. T., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review*, 121(2), 262-276.
<https://doi.org/10.1037/a0035940>
- Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., & Roediger, H. L. III. (2015). Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *American Psychologist*, 70(6), 515-526. <https://doi.org/10.1037/a0039510>
- 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*, 113(2), 304-309. <https://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(1), 10-65. <https://doi.org/10.1177/1529100616686966>

Yaniv, I., Yates, J. F., & Smith, J. K. (1991). Measures of discrimination skill in probabilistic judgment. *Psychological Bulletin*, 110(3), 611. <https://psycnet.apa.org/record/1992-04024-001>

Yarmey, A. D., Yarmey, M. J., & Yarmey, A. L. (1996). Accuracy of eyewitness identifications in show-ups and lineups. *Law and Human Behavior*, 20(4), 459-477. <https://doi.org/10.1007/BF01498981>

Yates, S. Q. (2017, January 6). Memorandum for heads of department law enforcement components, all department prosecutors. Washington, DC: Office of the Deputy Attorney General, U.S. Department of Justice. <https://www.justice.gov/file/923201/download>