Testing the Reflection Assumption: A Comparison of Eyewitness Ecology in the Laboratory and Criminal Cases

Heather D. Flowe, Anna Carline, and Nilda Karoğlu

Correspondence concerning this article should be addressed to Heather Flowe, University of Birmingham, School of Psychology, Hills Building, Edgbaston Park Road, Birmingham, B15 2TT, United Kingdom.

Contact: h.flowe@bham.ac.uk.

Author Note: We would like to thank the anonymous reviewers and the Editor, Dr Mark Coen, for providing invaluable feedback on the manuscript. We are also thankful to the many individuals who assisted us in coding the archival records and research literature, Jasdeep Chahal for assisting us in proofing an earlier draft of this paper and Professors Roy Malpass, Sigfried Sporer, Colin Tredoux, and the late Professor Ebbe Ebbesen for the many illuminating discussions with us on the topic of this manuscript.

Abstract

In the US, experts are often called upon to provide evidence during criminal trials regarding eye-witness identification research. A key factor relates to issue of probative value. To what extent are findings from the laboratory studies generalizable to the real-world? In order to answer this question, this article explores the issue of *eyewitness ecology*, a term referring to the environmental context in which people witness crimes, which includes characteristics of perpetrators and the viewing conditions, as well as the identification context. Specifically, we explore the extent to which the typical eyewitness ecology found in the laboratory studies reflects or is similar to real-world conditions. We coded the characteristics of the published literature on criminal identification in the laboratory (N = 309), and the results were compared to the characteristics of a stratified random sample of felony cases (N = 721) obtained from a large metropolitan district in the United States. This analysis demonstrated that in the criminal cases compared to the laboratory studies, duration of exposure to the culprit and retention interval length were significantly longer, and weapons, violence, and showup identifications were more prevalent. Additionally, the laboratory studies and criminal cases differed with respect to participant/witness race. These findings indicate a need to broaden the range of conditions employed in the laboratory to increase the applicability of eyewitness identification research to the legal system.

Keywords: eyewitness identification, eyewitness ecology, expert evidence, archival studies, field studies, lineup, estimator variables

Testing the Reflection Assumption:

A Comparison of Eyewitness Ecology in the Laboratory and the Field

In the US, experts in psychology and the law are often called upon to share their knowledge and understanding of the eyewitness identification literature with legal practitioners. For eyewitness experts, this often means testifying for the defence regarding the factors that can negatively impact eyewitness memory (e.g., Kassin et al., 2001; Kassin et al., 1989; Pezdek, 2009), or making policy recommendations for the handling and preservation of eyewitness testimony (e.g., Wells et al., 1998). Inherent in these applications is a *reflection assumption*: a term we use to refer to the presumption that the research conditions found in eyewitness identification research reflects the eyewitness context in "real-world" cases. Stated another way, individuals summarising how different factors affect eyewitness performance are assuming that the empirical findings, the vast majority of which come from laboratory studies, will generalise to real-world cases (e.g., Pezdek, 2009; Yarmey, 2001). However, whilst the reflection assumption may be reasonable, it has been given little systematic empirical treatment.

Many of the debates regarding the applicability of laboratory studies to the legal system have revolved around the question of generalisability. Some have argued that we are in a position to generalise because the laboratory methods used to study eyewitness identification are diverse enough to capture the essential characteristics of real-world crimes (e.g., Deffenbacher, 1984; Haber & Haber, 2000; Loftus, 1983; Pezdek, 2009; Yarmey, 1997; Yarmey, 2001), whereas others have cautioned that we must be sensitive to differences between the two settings with respect to the environmental and emotional context in which eyewitness identifications are carried out (e.g., Konečni & Ebbesen, 1986; Elliott, 1993; Flowe et al., 2009; Malpass & Devine, 1981; Pachella, 1986; Yuille, 1993). For example, Weiner and colleagues pointed out that some laboratory studies may oversimplify the visual behaviour of actual witnesses, as participant witnesses have their attention directed to the perpetrator, making generalisations difficult. Yuille and Wells (1991) have argued for the necessity for more comparisons between experimental research and field contexts so that the similarities and differences between the two can be enunciated. Although we would also want to know how memory performance between experimental research and field contexts differs, a more basic and unanswered question is: *How different are eyewitnessing conditions in the laboratory and in actual criminal cases?*

In this article, we examine *eyewitness ecology*, a term we coin here to refer to the context in which people witness crimes and take eyewitness identification tests, both in the laboratory and in real-world. The aim is to describe the ecological range that has been employed in laboratory research,[1] and compare it to those found in real-world cases. The primary reason to undertake this study is because the results from laboratory research are often applied to eyewitnesses in the real-world. First, a search of U.S. appellate cases on Lexis/Nexis database reveals that eyewitness experts testify in a wide range of different types of cases. When experts testify in these cases, the underlying assumption is that the eyewitness ecologies in the laboratory studies are similar enough to real-world eyewitness environments to warrant theoretical generalisation. Our focus on eyewitness ecology is inspired by the work of Egon Brunswik, who argued that researchers should particularise their experiments such that they are representative of the ecology, or habitat, to which generalisations are intended (Hammond & Stewart, 2001). Second, procedural recommendations[2] derived from laboratory research affect all cases in the legal system once they are implemented, not just potentially problematic cases in which the

---

[1] Throughout, the laboratory studies will be interchangeably referred to as ‘laboratory studies’ or ‘the studies’.
[2] Relating to, for example, best practice regarding the collection and preservation of eyewitness evidence.

witnessing conditions are poor. Accordingly, it is important to compare the range of eyewitness ecologies in studies with those found in real-world cases and assess the extent to which they overlap. In the main, we are undertaking what Malpass and colleagues (2008) referred to as a *study space analysis*, wherein individual factors in each of the studies (e.g., independent variables, dependent variables, methodological and procedural strategies) are defined in a matrix. Subsequently, the individual study matrices are merged so that each entry identifies the number of studies that utilised the corresponding characteristics.

The first objective of the research was to analyse the overall distribution of the conditions found in the laboratory and 'archival cases': real-life felony cases that were referred for prosecution. Second, the ecological conditions of laboratory and archival cases were compared to determine whether there are areas in which additional research seems warranted. For instance, how long are eyewitnesses exposed to the culprit in the archives and in the laboratory? How often are eyewitnesses subjected to violence? How often do eyewitnesses interact with the perpetrator before the onset of the crime? What are the lower and upper limits for the retention interval between the crime and the identification test? Does the retention interval length vary depending on the type of identification procedure employed? Answers to such questions can offer some guidance to researchers who are interested in utilising real-world parameters to inform the types of procedures and methods they use to study eyewitness identification in the laboratory. Our study space analysis revealed differences between the real-life cases and laboratory studies, particularly with respect to the race of the participant/witness, the duration of exposure to the culprit and retention interval length, and the use of weapons, violence, and showup identifications.

Methodology

Laboratory Studies

Sampling. A comprehensive literature search was performed using the Psychological Abstracts Information Services (PsycINFO) and PsychARTICLES. PsychINFO covers publications from almost 2,500 scholarly peer reviewed journals from more than 50 countries and PsychARTICLES covers 110 journals, including all APA published journals and Canadian Psychological Association journals. Our search covered the years 1597-2014, which are dates that coincide with a recent National Academy of Sciences report on eyewitness research (National Research Council, 2016). A selection of key words and title word searches related to our project aim (e.g., eyewitness, lineup) were entered as appropriate. Our only criteria for inclusion were that study participants were presented with a lineup or a showup task. We chose to include only published laboratory studies, because courts, in determining reliability, may consider whether the scientific evidence and methodology has been tested and subjected to peer review or publication and generally accepted by the research community (United States v. Williams (583 F.2d 1194 (2nd Cir. 1978), United States v. Downing (753 F.2d 1224 (3rd Cir. 1984), Daubert v. Merrell Dow Pharmaceuticals, Inc. (113 S.Ct. 2786 (1993)). From this search, the features of 444 studies from 309 peer-reviewed published papers, involving 93,189 participants were coded (218 papers included 1 experiment, 63 papers included 2 studies, and 28 papers included 3 or more studies). The journals in which these studies most frequently appeared were as follows: 82 studies (18%) were published in '*Law and Human Behavior*', 77 (17%) in the '*Journal of Applied Psychology*', and 72 (16%) in '*Applied Cognitive Psychology*'.[3]

[3] Due to space limitations, it is not possible to cite all the laboratory studies included in the analysis. Please contact the corresponding author for a complete list.

Of these selected studies, 53% presented the target (i.e., the person being observed, such as the culprit) in a video, 13% live in a laboratory, 10% in a photograph, 9% live in an auditorium or a classroom, 9% live in a natural setting, 5% in a slide sequence, and 1% in a composite drawings or audio tape. The target was portrayed under conditions that were not criminal in 38% of the studies (e.g., laboratory participants memorized a photograph, or field participants were asked to identify a customer or a researcher with whom they had previously interacted). In the remaining studies the target committed theft in 35%, robbery in 14%, other types of crimes in 4%, a bomb plot in 3%, assault in 2%, burglary in 2%, and vandalism in 1% of the studies.

Coding Scheme. We recorded the number and type of independent variables investigated, whether the independent variables were correlational, how the variables were manipulated (between or within subjects) and the setting in which the experiment took place (laboratory, home, store, or other natural setting). In addition, the number and type of participants recruited and participant demographics were coded as the primary “eyewitness” variables. The “perpetrator” characteristics gathered from the laboratory studies included the number of targets shown, target physical features, and whether the target was disguised. For the witnessing condition variables, we coded whether participants were “victimized”, exposed to a weapon, witnessed the crime alone or in a group, and whether participants knew they were in a memory experiment prior to seeing the event. The critical event (i.e., the simulated crime, or other circumstances in which the target was portrayed) and target exposure duration (in minutes) were also recorded when available. Finally, characteristics of the identification task that we coded were: whether participants were asked to describe the target, type of identification task administered, the actual and functional size of the lineup, method of lineup presentation (simultaneous or sequential), whether the target was present, how foils were chosen (with or without piloting), who constructed the lineup (experimenter, local police or other method), and the description and identification retention intervals.

In circumstances where certain information was not provided, we did not speculate as to the nature of the missing data values. For congruency with the archival analysis, data regarding "culprit" characteristics were based on the total number of simulated perpetrators, and data pertaining to "eyewitnesses" was analysed based on the total number of simulated eyewitnesses.

Archival Cases

Sampling. The characteristics of the laboratory studies were compared to a stratified random sample of 721 felony cases (robbery $n = 238$, rape $n = 301$, assault $n = 182$) that were referred for prosecution, by the San Diego Police Department, to the San Diego District Attorney's Office between January 1, 1991 and December 31, 1995. Of these cases, 65% ($n = 468$) were accepted for prosecution ($n = 208$, rape; $n = 176$, robbery; $n = 84$ assault), whilst in the others the issuing District Attorney dropped the charges against the defendant.[4] These three specific types of crime were selected because they were serious felony offenses that were most likely to involve eyewitness testimony as a pivotal factor in each case. In these cases, a suspect had been arrested by the police, and the case had been handed over to the prosecutor's office to determine whether charges should be filed. Both rejected and accepted cases were included in the sample because we wanted to include a wide range of eyewitness ecologies. On one end of the spectrum, the rejected cases might be largely representative of cases in which the eyewitness

[4] Recommended guidelines for the collection of eyewitness evidence impact police handling of all cases, regardless of whether the D.A. issues charges. Therefore, any attempt to determine how well laboratory research conditions generalise to real crime situations should also include cases that do not reach the prosecution stage.

ecology was likely to lead to poor memory formation, whereas on the other end of the spectrum, the accepted cases might be largely representative of circumstances in which the eyewitness ecology promoted stronger memory representations.

The sample represents 1,319 eyewitness testimonies (403 of the witnesses were from robbery cases, 450 from rape cases and 466 from assault cases). The number of witnesses in each case ranged from 1-10, but typically the cases involved the testimony of 1 eyewitness (Robbery: 40% involved 1 eyewitness, range = 1-9; Rape: 75% involved 1 eyewitness; range = 1-10; Assault: 42% involved 1 eyewitness, range = 1-10).

Coding Scheme. In line with the laboratory studies, case files were coded both at the defendant level and the individual eyewitness level. With the exception of the exposure duration variables (crime and target), lineup functional size, and weapon presence and use, the coding procedures previously described were utilized. As for the exposure duration variables, because this information was not actually available, estimates of these time intervals were made when coders felt they could reasonably approximate it from the eyewitness statements (if any) described in the police crime report. Coders estimated lineup functional size if a copy of the photographic lineup was available in the case file. Functional size was estimated based on the number of persons in the lineup that fitted the physical description of the culprit. Finally, for weapon presence and use, an eyewitness' awareness of the threat or use of a weapon, rather than the actual presence and use of weapon, was coded from the archival case files.

For descriptive purposes, and to assist with the design of future experiments, a few additional variables were coded in the archival cases. These included viewing conditions (i.e., lighting and distance away from the perpetrator), whether the police reported that the witness was under the influence of alcohol and/or drugs, and distinctive physical features of the perpetrator noted by the police in the report at the time of the arrest. These variables were not coded in the experimental studies because the information was usually not provided in the papers.

For the purposes of this study, a "case" was defined by a single defendant, and in the event of a file involving felony complaints against more than one individual, the file was seen as involving more than one case and was coded accordingly. As defendants were often charged with more than one crime, cases were assigned to a crime category (rape, robbery, or assault) based on the most serious charge reported on the felony complaint. The variables coded were taken from a number of sources in the case files, such as the crime report, arrest report, probation officer’s report, and the investigator’s follow-up report. Due to the large volume of case files, the completeness and content of the case files varied from case to case (case files ranged in size from a single file folder to several large boxes). As a result of this, the number of responses varied across each variable.

Coder Reliability. Before coding information from the criminal case files, the 46 coders involved in this portion of the study were provided with detailed written and verbal instructions, along with a number of practice cases to code. Intercoder reliability was measured by assigning one randomly chosen case to each of the coders at some point during the data collection process and monitoring intercoder agreement for the variables in this case. Given the extremely basic level of coding involved in this study, the agreement between the coders was extremely high. To measure intercoder reliability for variables that required estimations on the part of the coder, Cronbach's alpha was computed to examine reliability for each variable (see Howell, 2002). The values obtained for these alphas were .86 for exposure duration, .98 for description latency, and .89 for identification latency.

Data Analysis. Comparisons between the laboratory studies and the archival cases were made at both the defendant and the eyewitness level. For the categorical variables, chi-square analyses were performed to determine the statistical relationship between the variables manipulated by researchers, and those naturally occurring in the real-world. Since the retention interval and duration of exposure data were not normally distributed, the Mann-Whitney U statistic was used to compare the archives and laboratory studies on these variables (see Howell, 2002).

Archival results were expected to differ for some variables depending on the type of crime committed (e.g., the majority of witnesses in rape cases are female). For these variables, separate analyses were conducted for each crime category (rape, robbery and assault).

Exploring the Differences in Eyewitness Ecologies

In this section we provide an overview of the differences in the eyewitness ecology between the real-life cases and laboratory studies. The analysis is structured around the following elements: independent variables, eyewitness, perpetrator, witnessing conditions, and suspect identification. Thereafter, in the discussion section, we develop a more detailed analysis of the major differences pertaining to race, the duration of exposure to the culprit and retention interval length, and the use of weapons, violence, and showup identifications.

Independent Variables

Of the 444 studies, 77% employed a true experimental design, 20% included both experimental variables and correlational variables (such as participant age or personality traits), and the remaining 3% exclusively examined correlational variables. A one-way chi square analysis indicated that this distribution of research designs departs from equiprobability chance expectation, χ^2 (2, N = 444) = 400.96, $p < .0001$. The laboratory studies typically manipulated 2 (mode) independent variables (M = 2.55, SD = 1.44, range = 1 to 14 independent variables). Table 1 displays the top ten independent variables found in the research literature. For every variable listed, except for "memory ability" (which was most often operationalized as a continuous variable), the mode number of design levels was 2.

Eyewitnesses

Gender. Participant gender was available for coding in 51% (n = 219) of the studies and for 95% (n = 1,251) of the archival witnesses. Across all crime categories, 43% of the research participants and 46% of the archival witnesses were male, a difference that is not statistically significant, χ^2 (1, N = 51,906) = 3.70, p = .05. Gender composition in the archives varied by case type. Most real-world witnesses were male in robbery (65%) and assault (60%) cases, whereas the majority of the witnesses in rape (84%) cases were female, with the association between gender and case type being statistically significant, χ^2 (2, N = 1470) = 761.61, $p <$ .0001.

Age. The majority of the laboratory studies recruited college students to serve as witnesses (n = 296), and other adult populations (n = 51). Participants under the age of 18 (n = 25) were recruited less frequently (65 studies drew subjects from multiple populations and 7 studies did not report any sampling information). Across research participants, 69% were college students, 20% were from other adult populations, 10% were children and 1% were adolescents. The age of the participants ranged from 2 to 94 years (101 studies reported this information). For archival eyewitnesses (N = 1,133), the age range was 4 to 86 years, with the average age being

29.28 years (*SD* = 13.04, median = 26). The age distributions for the archives and laboratory studies were not statistically compared owing to differences across them in how age information was recorded.

Race. Race was known for 26,806 research participants (81 studies reported participant race) and for 1,142 archival eyewitnesses. The distribution of eyewitness race in the archives significantly differed from the distribution found in the laboratory studies, χ^2 (4, N = 26120) = 2103.81, $p < .0001$. Whereas 77% of laboratory witnesses were White, only 49% of real-world eyewitnesses were white.[53]As for other racial categories, the background of participants in the studies was: 3% Black, 7% Asian, 3% Hispanic, and 10% were designated as belonging to other racial categories. In the archives, the reported racial categories were: 18% Black, 6% Asian, 26% Hispanic, and less than 1% were reported to identify with other racial categories. The distribution of these other racial categories significantly differed across the archive and real-world cases, χ^2 (3, N = 6746) = 943.11, $p < .0001$. In the real-world cases, 52% of the identification attempts in which the race of the eyewitness and suspect was known (N = 211) were cross-race. As for the total of eyewitnesses, 42% of the witnesses were cross-race and this information was reported for 1073 eyewitness. Only 60 laboratory studies reported both the race of the suspect and eyewitnesses simultaneously, and 63% of these (38 studies) employed either cross-race target or participants.

Relationship to Target. The archival witnesses were acquainted with the suspect in 92% of the assaults, in 79% of the rapes, and in 21% of the robberies. However, in the laboratory studies none of the participants were acquainted with the target. In a few of the studies,

[5] In describing the racial backgrounds of research participants, eyewitnesses, and defendants, we adopted the racial nomenclature that legal officials used in the case files.

researchers indicated that when subjects volunteered to participate and reported to the experimenter that they were familiar with some members of the lineup, they were immediately excluded from the data sample.

Alcohol and Drug Use. A total of 13% of all real-world witnesses were under the influence of drugs (n = 28), alcohol (n = 122) or a combination of the two (n = 26) at the time the crime was committed. Most of these witnesses (73%) were witnesses from the rape cases. A description of the culprit was given to the police by 45% of witnesses who were under the influence of drugs and/or alcohol (n = 79). Additionally, 22% of the intoxicated witnesses were provided with an opportunity to identify a suspect (n = 39); 44% of these cases involved defendants and witnesses previously unacquainted. Few eyewitness papers involved intoxicated participants (Harvey, Kneller, & Campbell, 2013; Yuille & Tollestrup, 1990; Yuille, Tollestrup, Marxsen, Porter, & Herve, 1998).

Perpetrator

Gender. The gender of the targets significantly varied in the archives compared to the laboratory studies, χ^2 (1, N = 1492) = 89.48, p < .001. While both laboratory and actual criminal cases involved primarily male targets, the laboratory studies utilized a greater number of women (n =215) compared to the archives (n = 41). When women were employed as the target in the studies, 52% were portrayed in conditions that were noncriminal, 42% engaged in simulated thefts, 5% in burglary, and 1% in robbery scenarios. In relation to the archives, 64% of the female defendants were alleged to have committed robbery, 32% assault and 4% rape. For theft and robbery crimes, the proportion of female perpetrators/defendants did not differ significantly across the archive and laboratory studies (14% versus 7%, respectively), χ^2 (1, N = 239) = 1.92, p = .16.

Age. Information was provided in 151 of the laboratory studies regarding the general age range of the culprit (rarely was the exact age of the culprit provided). In most studies, a college student or an adult under the age of 26 was the target (85%), followed by adults over the age of 26 (14%). A child was the target in 1% of the studies. In the archives, the defendants were on average aged 29 years (SD = 9.10, range = 16 to 79, median = 28 years, n = 628). The age of the culprit/defendant could not be compared across the archive and laboratory studies owing to differences between the two in how this information was reported.

Race. The distribution of target race (N = 620) in the archives significantly differed from the race distribution for targets found in the laboratory studies (N = 397, with 42% of the studies reporting this information), χ^2 (4, N = 1017) = 299.22, p < .001. In the archives, 30% of the defendants were White, whereas 85% of the targets in the studies were White. With regard to the other racial categories found for the targets in the studies, 8% were Black, 2% were Asian, 3% were Hispanic, and less than 2% categorized as "other". In the archives, the distribution was 33% Black, 3% Asian, 31% Hispanic, and 3% were categorized as "other". The archive and laboratory studies with respect to these other racial categories significantly varied, χ^2 (3, N = 495) = 23.42, p < .0001.

Disguises. The description(s) of the culprit given by the eyewitnesses included a disguise in 6% of the archival cases (n = 44). In most cases the items reported to disguise the perpetrator's appearance most frequently were caps and hats (68%), followed by glasses (20%), bandanas (7%) and stockings (5%). In the laboratory studies, perpetrators less frequently wore a disguise compared to the archives, χ^2 (1, N = 1694) = 9.22, p < .00001. Disguises were used by simulated culprits in 5% of the studies (n = 25); with hats (36%), different clothing (12%), hairstyle (12%), headscarf (8%), glasses (8%), stocking (4%) a mask (4%), and sunglasses (4%), serving as the disguises. In 12% of the studies, no information was provided about the type of disguises.

Distinctive Features. Specifically distinctive physical features were noted in the police report for 33% of the defendants (n = 241). Of these defendants, 63% had one or more visible tattoos, 17% had prominent scars on their body, 11% had faces with tattoos, scars, or some other unusual feature, and 3% had missing, broken or gold-capped teeth. None of the papers in the literature mentioned whether their culprits had such features, though one study manipulated whether the target wore an Elvis wig or had a chipped tooth (Searcy et al., 2000). As such, no statistical comparisons were made.

Witnessing Conditions

Victimization. In the archives, 56% of the eyewitnesses were crime victims, whereas 7% of research subjects were led to believe that they were victims of a crime, a statistically significant difference, χ^2 (1, N = 94,436) = 4113.93, $p < .00001$. Throughout the laboratory studies, the victimizations were limited to theft due to ethical reasons. Real-world witnesses were victimized 70%, 55%, and 42%, of the time in rape, robbery and assault cases, respectively.

Critical Event. In the laboratory studies, 47% of the participants were aware that they were taking part in an experiment before viewing the target. In cases where participants were not aware, 43% were presented with a noncriminal event and 99% were debriefed before the identification test was administered. Across participants who were aware from the beginning that they were taking part in an experiment, 40% were not told that they would be given an identification test until after the critical event was presented. It was unclear from the archival records when witnesses became aware that they were in fact witnessing a crime; hence, no statistical comparisons were made in this regard.

As the locations in which the critical event occurred were so different for the archives compared to the laboratory studies, only descriptive information will be provided here. In 9% (n=41) of the studies, there was no mention of where participants viewed the critical event. Of studies which did state where the critical event took place, 71% (n = 287) reported a laboratory setting, 19% (n = 77) a lecture hall setting, 7% (n = 27) in public places (4 studies presented the event in multiple settings), 2% (n=7) in the police station or military training service, 1% (n=3) online via internet, in the participant's home in 1 study, and in the fire station in 1 study. The crime location was revealed in 655 of the archive cases with the following locations/settings being reported: at a residence (52%), on the street or in an alley (18%), in open public places (13%), in other outdoor settings (8%), in other indoor settings (3%), in multiple locations (3%) or in vehicles (1%).

Alone or in a group. For 394 laboratory studies and 257 of the archive cases, information was available regarding whether the witnesses observed the crime by themselves or with other witnesses. The critical stimulus was viewed autonomously by a participant more often in the studies compared to the archives (51% versus 18%), χ^2 (1, N =651) = 74.45, $p < .001$. Furthermore, in 222 of the archival cases it was clear that 2 or more witnesses were previously acquainted with one another. None of the research investigations reported whether witnesses were acquainted with one another prior to observing the event, and hence no statistical comparisons were made in this regard.

Event Exposure. Table 2 presents the event and target exposure durations for the studies and the archives by the type of event witnessed. For the archival cases in which event duration was estimable (*n* = 507), compared to the eyewitness studies (*n* = 335; 13 studies presented multiple events), event exposure time was considerably longer (Mann-Whitney U, z = -19.91, *p* < .0001) and the data were encompassed by a broader range of durations. For studies, the mean exposure time was 2.32 min (*SD* = 5.76, range = 0.03 to 40 min, and median = 1 min). For the archives, the criminal event lasted on average 59.87 minutes (*SD* = 390.75, range = .08 to 7200, and median = 8 min).

Target Exposure. As shown in Table 2, witnesses (*n* = 403) in the archives were exposed to the target for a longer duration compared to research participants (Mann-Whitney U, z = -16.42, *p* < .0001). Duration of exposure to the target was reported in 283 studies (4 studies involved multiple crime presentation); on average, research participants viewed the target for 1.80 min (*SD* = 5.58, range = 0.03 to 40 min, and median = 0.4 min). Archival witnesses viewed the culprit for 65.74 min on average[6] (*SD* = 205.17, range = 0.00 to 2423, and median = 7.8 min). 72% of real-world eyewitnesses viewed the suspect's face for 10 min or less and 44% of these witnesses saw the suspect for 3 minutes or less. This information was provided for 832 eyewitnesses.

Violence. Violence did not feature prominently in the laboratory studies, likely owing to ethical considerations; thus, no statistical comparisons across the laboratory and archives will be made with respect to violence. Only 2 studies manipulated the level of violence (Clifford & Hollin, 1981; Cutler et al., 1987). None of the published papers examined the effect of prior exposure to the culprit before violence. In the real-world cases, 59% of eyewitnesses were exposed to violence (this variable was known for 1,251 witnesses). Of the witnesses exposed to violent events, 117 first interacted with the culprit during the violent act, whereas 176 first

[6] That the mean duration of exposure to the culprit is longer compared to the mean crime duration is not unusual, as some witnesses viewed the culprit immediately before and after the crime.

interacted with the culprit before the violent act (for 46 witnesses, this variable was unknown). The remaining 399 eyewitnesses exposed to violence personally knew the culprit. Additionally, of witnesses who attempted to identify a stranger-culprit, 84% of those experiencing violence were between 0-36 inches away from the culprit, whereas only 44% of witnesses who were not subjected to violence were within this range of proximity.

Weapon Exposure. A much larger proportion of archival witnesses were exposed to weapons (54%) compared to research participants (9%), χ^2 (1, N = 94,508) = 3136.24, $p <$ .0001. A total of 37 staged crimes involved weapons. Of witnesses who saw a weapon, 49% viewed a gun, 32%, a bomb, 8% a knife, 3% a syringe and 3% a bottle, 3% a truck (one study did not specify what the weapon was). Archival witnesses (N=1,046) viewed a greater variety of weapons than research participants (37% reported the culprit's hands or feet were used as a weapon, 24% were exposed to a penis, 12% saw a firearm, 10% a knife, 7% a blunt instrument, 1% a ligature, and 9% fell into the other category, which included weapons such as rocks, bottles, and vehicles). Moreover, several archival witnesses (n = 286) saw more than one type of weapon, whereas this was not the case for any of the research witnesses. Duration of weapon exposure in the archives ranged from 3 seconds to 15 hours, with eyewitnesses typically exposed for 1 minute. In the studies, duration of weapon exposure was often not reported, hence no statistical comparisons between the laboratory studies and the archives could be made.

Suspect Identification

Description. In the studies (37%), subjects were asked to give a description of the targets. In the archives, a description was given by 500 witnesses, 746 did not give a description, and one witness died before the investigators could obtain a description (this variable was known for 1,247 witnesses). The proportion of participants who gave a description was significantly greater in the archives compared to the laboratory studies (37% versus 40%, respectively), χ^2 (1, N = 94,436) = 5.06, p = .02.

Identification Task Construction. Of course, the police arranged all the identifications in the archives. Unfortunately, the case files did not provide us with information on how the police selected the foils for the lineups. Therefore, statistical comparisons between the laboratory studies and the archives were not made, and only descriptive information will be provided. We will concentrate on how the lineups were constructed. Note that police are advised to construct lineups for a given target by selecting fillers (i.e., persons known by the police to be innocent of the crime) who resemble the witness' description of the perpetrator (ref), which research suggests is a fair way to construct a lineup (Technical Working Group for Eyewitness Evidence, 1999). Another approach that could be used is to select fillers based on how similar to the target/suspect they visually appear. The purpose of fillers is to decrease the odds that the witness identifies the suspect based on guessing alone. To serve this purpose, the fillers should not be too dissimilar in appearance to the target/suspect. Further, researchers and the police can pretest their lineup materials to ensure that they are fair and unbiased (see Tredoux, 1998). We focus on lineup construction in this section because it can play an important role in lineup identification accuracy, and thus, should be considered when generalising across laboratory and real world cases.

In 67% of the laboratory studies, the researchers selected fillers for the lineup because the researchers simply thought the fillers appeared similar enough to the target. In the studies in which lineups were pretested, the fillers were selected on the basis of similarity ratings gathered from subject raters, who rated the similarity of the target/suspect against each of the fillers. The fillers were selected by matching them to a description of the target in 35% of the studies, by matching them to the target's photograph in 21%, by using other strategies in 20%, and no information was provided as to what strategy the researchers used in 24% of the studies. The lineups were mock witness tested in 28% of the studies to determine whether the fillers were adequate. The researchers asked the police to construct the lineup in 5% of the studies.

Lineups size, or the total number of people who are in the lineup, varies across police jurisdictions, with 6 being the norm in the US and 9 being the norm in the UK, as examples. Lineup size in the laboratory studies ranged from 1 to 120, with 56% of the studies involving 6 persons. However, the functional size of a lineup is also important. Functional size refers to the number of lineup persons who are similar in appearance to the target/suspect. Nominally, a lineup may contain 6 persons; but, if 4 fillers are not similar to the target/suspect, the functional size of the lineup will be 2 persons. The odds of a witness being able to identify the target/suspect based on guessing alone are higher in a lineup that has a functional size of 2 as opposed to 6 persons. Functional size can be determined based on pretesting the lineups using mock witnesses. The functional size of the lineup was determined in 16% of the studies, and ranged from 1 to infinity (in some studies, lineup member similarity was manipulated). Further, the position of the target/suspect in the lineup was reported in 49% of the studies. For a lineup in which the photographs are presented simultaneously to the witness, the photographs are usually arranged in a 3 x 2 array, with the first position being the leftmost photo in the top row, the second positon being the next photograph on the right, etc. It is important to ensure that the target/suspect does not always appear in the same position for every participant in case people are biased to select photos that appear in a certain position (e.g., the first or last photograph; see x for example). For the studies in which the positon of the target/suspect was provided, 59% counterbalanced or randomized the position of the target, 22% used a single position and 19%

utilized more than one position (typically 2 positions). Typically, the target was placed in positions 2 or 4 (27% of the studies used one of these positions).

In the archives, "lineup size" or the number of persons presented in a given identification procedure ranged from 1 to 50, with 51% of all eyewitnesses presented with 1 person for identification. Thus, witnesses were often presented with live showups. This may be because witnesses are most likely to be asked to attempt an identification of the perpetrator when a suspect is immediately apprehended at the scene of the crime. Researchers and the courts have long considered showups suggestive; nevertheless, the police present live showups when public safety and the liberty of the suspect who has been detained are at issue (see Gonzalez, Ellsworth, & Pembroke, 1993 for a review of the legal issues and as well as a laboratory and archival study of showups versus lineups). In the present study, when photographic lineups were organised, 6 persons were presented for identification (range = 1 to 9 persons) in 74% of the procedures. The functional size of the photographic lineups ranged from 1 to 9 persons. For 64% of the 6-person lineups, the functional size was estimated as 6 persons by the coders. The rate at which the target was placed in positions 1-6 was as follows: 10%, 31%, 21%, 17%, 16%, and 4%.

Identification Procedure. As the methods used to identify suspects were so markedly different in the laboratory compared to the archives, no statistical analyses were conducted. In the archives, 31% of the witnesses (n = 406) were asked to identify the suspect, and lineup tests were presented significantly more often to victims compared to eyewitness bystanders (67% versus 33%) χ^2 (1, N = 406) = 46.22, $p < .0001$. Of those presented with a lineup, 344 positive identifications of the suspect were made.[7] Additionally, approximately 1% of the eyewitnesses in

[7] Only the characteristics of the first identification test for each eyewitness were included in our analyses. We found that 25 eyewitnesses in the archives (6%) were presented with more than one opportunity to identify the suspect.

the archives who were given an identification test participated in the creation of a composite drawing, and none of the witnesses viewed mugshots prior to being given an identification test. Approximately 5% of the research participants viewed mugshots or a composite drawing before being given the identification test. Witnesses who were given an identification opportunity typically underwent one identification procedure (range: 1 to 2 opportunities).

As displayed in Table 3, while live showups were most often presented to real-world witnesses (51%), photographic lineups were most often presented to research participants (78%). Live showups were viewed by less than 1% of research participants, and no real-world witnesses viewed video lineups or listened to voice lineups. The lineup faces were viewed simultaneously by 74% of research participants, sequentially by 23%, 2% of participants were presented with showups and 1% with an elimination method. The researchers manipulated whether the target was present in the lineup for 67% of the participants. 16% of participants were shown only target present lineups, 9% of participants were shown only target absent lineups, and 8% of participants viewed both target present and target absent lineups.

Table 4 presents identification outcomes in the archives by the type of identification procedure that was conducted and the relationship between the eyewitness and perpetrator. As demonstrated, most identifications in the archival cases were conducted via a live showup procedure. Additionally, 71% of witnesses asked to take identification tests were not acquainted with the perpetrator.

Retention Intervals. Descriptive results are given in Figure 5. The majority of descriptions were made on the same day as the incident in both the archival cases (97%) and the studies (87%). The length of time between the crime and when participants were tested could be computed for 23,499 research participants (143 studies provided this information) and 139 archival witnesses. As can be seen from Figure 5, real-world compared to archival witnesses experienced longer delays between the crime and giving a description of the culprit. Real world witnesses also had a longer delay between the crime and when were given an identification opportunity. Furthermore, 105 studies did not provide the exact length of time between the crime and when participants were tested. In these studies, participants viewed the target, completed a filler task, and then were given the identification test in 12 of these studies. Participants viewed the target, described the target, and then completed the identification test in 45 studies. Thus, the delay between the crime and when participants were tested was likely quite minimal. In 48 of the studies, no information was given as to when the identification took place.

Witnesses in real-world cases could potentially be asked to identify the culprit on occasions that follow their initial identification. To aid researchers in determining retention interval lengths in future studies, descriptive information about these occasions is as follows: the preliminary hearing typically occurred 30.52 days (SD = 29.77, range = 2.72 to 168.92, median = 17.65 days, n = 106) after the crime, and the trial period typically occurred 149.45 days after the crime (SD = 90.95, range = 30.87 to 404.81, median = 103.06 days, n = 24). Of the 796 witnesses involved in cases accepted for prosecution, 51% testified at the preliminary hearing. In the 24 cases that went to trial, 97% of the witnesses testified. Whether identification accuracy would be at the floor (i.e., very poor) and what factors may affect memory retention with delays as long as the ones we found in the archives remain open research questions.

Discussion

Our study space analysis revealed major differences between the police cases and laboratory studies and quantified the extent to which the characteristics that are related to identification accuracy have been studied in both archival cases and laboratory studies. For example, the archival analysis indicated that real-world eyewitnesses were more often male in robbery and assault cases, and more often female in rape cases. The average age of the eyewitnesses was 30 years, and half of the witnesses were White. Except in the robbery cases, most of the eyewitnesses had some degree of familiarity with the suspect prior to the crime. Most of the time eyewitnesses observed the criminal event in a group of two or more persons. More than half of the eyewitnesses interacted with the culprit before violence was involved. Further, in keeping with Horry, Halford, Brewer, Milne, and Bull's (2014) archival analysis, we found more than half of eyewitnesses were victimized and exposed to a weapon. With respect to the culprits in the real-world cases, the majority were male and a third of the culprits were White. Approximately a third of the culprits had distinguishing features, such as facial tattoos and gold-capped teeth. About a third of the witnesses were given an identification test, and most often, those who were given the test were victims rather than bystanders. The identification procedure most often used was a live showup procedure. In a live showup, the witness is shown the suspect live in person and asked whether this is the person who perpetrated the crime. In cases where a lineup was conducted, most often the suspect was placed in position 2.

While there was some difference between the real and laboratory world for almost every characteristic we compared (the exceptions were the age and gender of the target and eyewitness), the discussion that follows will largely focus on the variables about which experts testify in the courtroom as being strongly related to identification accuracy (Kassin et al., 2001;

Pezdek, 2009): cross-race identifications, weapon exposure, exposure to violence, culprit exposure duration, retention interval to identification, and the type of identification task.

Meta-analyses of the own-race bias find that cross-race manipulations significantly decrease suspect identifications and increase mistaken identifications in the laboratory (Meissner & Brigham, 2001; Shapiro & Penrod, 1986). Own-race bias is influenced by both perceptual expertise, where people have greater expertise with faces of their own-race than of other races, and in-group/out-group categorisation. Own race faces may be learned in a different manner than other race faces (see Ryder et al., 2015 for a review). Some evidence indicates that different types and depth of cognitive processing are partly responsible for the underlying mechanism of own-race bias. For example, own-race faces are found to be processed more holistically (Michel et al., 2006; Tanaka et al., 2004), processed more configurally (i.e., the interrelations of salient facial features are processed more than individual facial features) (Hancock & Rhodes, 2008; Megreya et al., 2011) and encoded at a deeper level than other-race faces (MacLin et al., 2004). The results from different studies also suggest that own-race bias is reliable across most races (e.g., Goldstein & Chance, 1980; MacLin et al., 2001; Megreya, et al., 2011). Moreover, own-race bias influences identification accuracy in the laboratory studies to the extent that eyewitnesses are 1.40 times more likely to make correct identifications when the perpetrator is own-race, but more likely to make misidentification when the perpetrator is of another race (Meissner & Brigham, 2001). Despite the evidence supporting own-race bias, field studies have reported inconsistent findings, with some researchers finding an association between cross-race eyewitnesses and suspects and identification outcomes (Behrman & Davey, 2001; Horry et al., 2012) and others not finding an association (Pike et al., 2002; Valentine et al., 2003; Wright & McDaid, 1996). In the present study, we found almost half of the real-world eyewitnesses (42%)

and more than three quarters of research participants (i.e., 76%) were of a different race than the culprit. However, it is important to note that the laboratory data on which this finding is based were drawn from a small number of studies (i.e., 60). This was because the majority of laboratory studies (i.e., 384) did not report the race of the perpetrator and research participants**.** Given the potential importance of this variable and its high occurrence in the real-world, more research is needed to manipulate cross-race eyewitnesses. On the other hand, if it is not feasible for the researcher to examine own race bias (e.g., the researcher has access to only White participants), researchers should consider designing studies where there are at least two perpetrator conditions, wherein the perpetrator is either White or non-White. Through conducting such studies, an opportunity will arise to determine to some extent, whether a given phenomenon of interest generalises for White participants across perpetrators. For example, Lawson and Dysart (2014) found cross-race effect in lineups but not in showup identifications.

Exposure to a weapon can reduce recognition accuracy in laboratory studies, as reported by Steblay (1992) in her meta-analysis of the phenomenon. Importantly, in contrast to archival studies that found no weapon-focus effect (e.g., Behrman & Davey, 2001; Valentine et al., 2003), a recent meta-analysis conducted by Fawcett et al., (2013) indicates that the presence of a weapon can decrease identification accuracy in laboratory studies and in real-world environments, depending on context. For example, the weapon focus effect is less likely to occur with relatively brief or relatively long durations of weapon exposure. The meta-analysis also found that the weapon focus effect was smaller when the retention interval between the critical event and identification test was relatively long. Our comparison of real-world to simulated cases shows that weapons are more likely to be threatened and used in actual crimes. While a range of weapons were threatened against eyewitnesses in the real-world, guns were the most frequently threatened (49%) in the pre-recorded simulations. In contrast, a perpetrator's hands and feet were the most frequently used weapons in the archives, with the use of a gun reported by approximately 1 in 10 eyewitnesses who were exposed to violence, a finding which is in line with the archival analysis conducted by Valentine et al., (2003). These results suggest additional avenues for laboratory research in which the effect of memory strength on identification accuracy is examined in relation to a range of different types of weapons including hands and feet. Additionally, given the potential importance of this variable, more research is needed to investigate how different types of weapons capture attention and whether or not weapon effect is exacerbated with the use of different weapons (e.g., gun vs. knife). Likewise, further research should examine the relationship between identification accuracy and weapon presence in relation to other factors that might reverse or eliminate the weapon focus effect (e.g., exposure to the perpetrator before he brandishes a weapon). Given the prevalence of weapons in real-world crimes, it is important to establish the boundary conditions of the effect. For example, research could investigate whether the size of the weapon focus effect varies depending on whether witnesses are exposed to the culprit prior to weapon exposure, or depending on the size and features of different types of weapons.

Most people are likely to experience heightened stress when witnessing a crime, especially a violent crime. In a recent meta-analysis of laboratory studies, Deffenbacher et al., (2004) found that face recognition is less accurate under conditions of increased stress. In contrast, Wright and McDaid (1996), Horry et al., (2014) did not find an association between eyewitness exposure to violence and suspect identification rates in archival cases. There are numerous factors that possibility interact with violence exposure and that qualify whether it has an effect on memory accuracy. Perhaps for this reason expert consensus is relatively non-existent regarding the reliability of the relationship between stress and eyewitness memory (Kassin et al., 2001). For instance, we found that witnesses in the archives who experienced greater levels of violence also viewed the culprit from a shorter distance than those who were not exposed to violence, a finding that has also been demonstrated in a previous field study and an archival analysis by Yuille and Cutshall (1986) and Horry et al., (2014), respectively. In addition, Horry et al., (2014) found an association between heightened stress and longer exposure durations. These results suggest that encoding conditions that promote greater memory strength (e.g., a relatively short viewing distance) could attenuate the negative effect of heightened stress on memory (see Horry et al., 2014). Our analysis of the laboratory studies reveals that further eyewitness research examining the moderators of the relationship between stress and identification accuracy are needed.

We found that the range of target exposure durations in the laboratory and in the field significantly differed. Participants who are shown the perpetrator for a longer length of time are more likely to make accurate identifications according to meta-analyses of laboratory studies (Deffenbacher et al.,, 2008; Bornstein, Deffenbacher, Penrod, & McGorty, 2012), and with increased suspect identifications in archival investigations (Horry et al., 2014; Valentine et al., 2003). Whereas the length of time that the culprit was in view was less than 1 minute for only 3 in 20 real eyewitnesses, this was the mode for subject eyewitnesses, with almost 4 in 5 given this limited viewing time. In addition, we found that archival witnesses viewed the culprit for approximately 133 minutes in rape cases, 16 minutes in robbery cases, and 10 minutes in assault cases. In contrast, in the laboratory studies, participants typically viewed the perpetrator for approximately 1.13 minutes in rape scenarios, 0.82 minutes in assault cases, and 0.54 minutes in robbery cases. To better understand the boundary conditions of the effects we study, we need to increase the duration of exposure in our laboratory studies. For example, are we less likely to obtain the own-race bias and weapon-focus effects as duration of exposure to the culprit increases? We strongly believe that eyewitness researchers should more routinely vary duration of exposure to the culprit as well as the length of the retention interval between the crime and the identification as a matter of course, no matter what the phenomenon of interest that they are investigating (e.g., own-race bias, weapon-focus effect), to assess whether the effect generalises across conditions.

Further, our findings showed that the delay between the crime and when a photographic lineup test was administered was significantly shorter in the laboratory compared to the archives. While laboratory witnesses typically waited approximately 20 minutes to identify the suspect from a lineup, archival witnesses waited approximately 11 days. However, while these longer retention intervals found in the real-world might raise reasonable concerns regarding identification accuracy, archival witnesses were likely to be questioned about the incident within minutes of the crime. Turtle and Yuille (1994) raised the issue that the way in which real witnesses experience retention intervals might differ from laboratory witnesses. As a result, typical memory loss functions found in the laboratory might be different for real-world witnesses. In support of this argument, Ebbesen and Rienick (1998) found that the accuracy of recalled events did not decay over time when participants were repeatedly questioned.

The general consensus among eyewitness experts (74% agreement) is that the use of showups as an identification task increases the risk of misidentification (Kassin et al., 2001). In a recent meta-analysis of photographic showup procedures, Steblay et al., (2003) found that showups yield equal hit rates. However, participants were more likely to not make an identification at all compared to a lineup. The latter result was qualified by suspect-perpetrator similarity, as high similarity innocent suspects were more likely to be falsely identified from a showup. We found that *live* showups are the most frequently used identification task in the real-world, presented to half of all eyewitnesses that were given an identification test. In contrast, the photographic lineup is most frequently used in our laboratory simulations, presented to 78% of subject witnesses. Only 3 experiments in the literature (Gonzalez et al., 1993; Valentine et al., 2012; Yarmey et al., 1996) compared eyewitness accuracy in live showups to lineups. Additionally, like Behrman and Davey (2001), we found that positive identifications of the suspect were significantly higher in showups compared to lineups. Additionally, in our archival sample, showups were typically administered within an hour of the crime, whereas lineups were conducted approximately 11 days later. These findings suggest that additional research is needed to examine the effects of live showup procedures under varying conditions on identification accuracy. This is potentially fertile ground for researchers who are interested in memory, as in a live showup there are cues, such as suspect gait and demeanour, which might affect memory retrieval in a manner that is different from a photographic showup.

While we feel that more attempts to reflect a greater range of real-world conditions would enhance the generalizability of laboratory studies, we acknowledge that not all conditions are equally feasible or ethical. Further, if research methods in a literature are similar in terms of the systematic errors that arise, researchers should select different methods to reduce such errors. This would allow for better capturing the effects of the phenomenon of interest and test research hypotheses (Blankertz, 1998). The onus for the researcher to do so is amplified when the research will be applied in the courtroom (Weiner et al., 2002). As examples, Maass and Köhnken (1989), and Morgan et al. (2004) have employed creative (and still ethical) procedures to create a potentially more stressful eyewitnessing situations.[8] Similarly, we are not suggesting that all variability in conditions be removed from empirical manipulations or that better reflecting conditions of the real-world will address all concerns regarding the external validity of laboratory research. However, we do feel that in addition to utilizing multiple methodologies, such as laboratory studies, archival research, and studies with actual eyewitnesses, more realistic eyewitness simulations will strengthen our understanding of what factors influence the accuracy of eyewitness identifications and make us less likely to over-generalise empirical results to real-world cases. In order to advance theory in the eyewitness memory domain, we need research that is high in experimental realism (i.e., studies that use controlled and artificial experimental tasks that nevertheless engross participants and elicit the psychological states of interest). We also need research that is high in mundane realism (i.e., tasks that are representative of situations that are like those that people might encounter outside of the laboratory) (Aronson, & Carlsmith, 1968). As an example of research that is high in experimental realism, one laboratory study we coded systematically examined the effects of distance and illumination on identification accuracy (Wagenaar & van der Schrier, 1996). The paper reported that accurate identifications were obtained from at least illumination of 15 lux and a viewing distance of no more than 15 metres. These results were confirmed by another study in which participants were exposed to faces of famous people or lookalikes (De Jong et al., 2005). However, before this research can be used in the courtroom (which the authors of the illumination study thought possible) research high in mundane realism is needed to determine how durations longer than 12 seconds (which was the

[8] For example, the ‘weapon’ used in Mass and Köhnken (1989) was a syringe. In this study, half the participants were approached with a syringe and half with a pen of a similar size; and in both groups half expected to receive an injection. It was considered that the threat of a needle would produce a fear reaction on the basis that most people are in the least averse to injections, but that this would not be unethical given that injections are common place in experiments. In Morgan et al., (2004) in order to measure the effects of a highly stressful situation, the study was conducted with active-duty military personnel during military survival training.

exposure duration used in the illumination study) or how a live presentation of the target (instead of presenting a portrait, which was the case for the illumination study) might alter the functional relationship observed between distance, lighting, and identification accuracy. As the present study has shown, real world witnesses encounter perpetrators, in person, for longer than 12 seconds. Thus, before we can generalise the findings to real world witnesses who view a live perpetrator in person for a given length of time, it may be important to investigate the impact of these factors (live presentation of the perpetrator at the given length of time) on identification accuracy.

While the goal of much of psychology is to develop universal theories of human behaviour, just how context insensitive experiments in the eyewitness memory domain can be before they lose their real-world value is an empirical issue. For example, does the videotaped showing of a staged crime to participants capture the essence of what it means to be a real eyewitness? Simulating a criminal event might be less important than, say, trying to capture other essential aspects of witnessing. For instance, we might speculate that features such as high situational ambiguity (e.g., “Am I really being robbed?”), or the need to decide within seconds how to respond to the situation (e.g., ‘Should I give up my person or property?’ ‘Should I fight back or run?’), are factors that are more important in terms of simulating the psychology of eyewitness identification than, say, staging for our subjects an event wherein a criminal activity is perpetrated against someone else. Finally, while archival studies examining eyewitness behaviour in actual criminal cases contribute to our understanding of eyewitness identification by allowing us to describe its features in natural environments, archival studies do not of course allow for a proper test of theoretical relationships between a factor and memory accuracy (Horry et al., 2014). Witnesses are not randomly assigned to conditions and numerous interrelated variables may have interactive effects on memory cannot be separated; thus, it difficult, if not impossible, to ascertain from an archival analysis how a given factor influences memory. Consequently, we need to be mindful of the real-world conditions to which we wish to generalise our results, and design our experiments so that they best reflect the generalisation conditions.

Whether we are appropriately simulating witnessing is a reasonable question to ask, for unlike the real-world cases, experimenters do not typically ask their participants how closely they paid attention, or screen out people who did not pay attention to or remember the crime (see Yuille & Cutshall, 1986, for an interesting discussion relating to how investigators might screen out witnesses who had a poor vantage point). To illustrate, Kurosawa (1996) used a typical staged crime scenario in a lecture hall setting and found that 33% of the participants could not recall seeing the staged incident. Additionally, 40-50% of the students were not able to say what the intruder did. Similarly, a study presenting a staged theft in a lecture hall setting reported that only 34 out of 147 students indicated they had witnessed the mock crime (Riske et al., 2000). While the police may not extensively question or ask the witness to make an identification if they did not attend to the crime, in “live” staged crime experiments, the results are rarely conditioned on such factors. Additionally, these results from Kurosawa (1996) and Riske et al. (2000) suggest that our crime simulations might not be as impactful as real criminal events; to our recollection, we coded only 3 studies (Hollien et al., 1983; Leippe et al., 1978; Hosch & Bothwell, 1990) reporting that subject witnesses to a staged crime attempted to intervene on behalf of the victim. While for obvious ethical reasons we cannot expose our participants to real criminal activity, studies such as these suggest that we should generalise laboratory results more cautiously, and that we should take seriously the issue of whether our simulations allow participants to be representative of real-world witnesses.

Overall, it is important that researchers are mindful of the ‘moderate and significant inverse relationship between research design and outcome’ (Welsh et al., 2010: 193; Weisburd et al., 2001). Accordingly, weaker research designs – particularly those which involve nonrandomized samples - are more likely to be prone to selection bias. Given the over representation of psychology university students in the laboratory studies, greater consideration should also be given to the diversity of the research participants to ensure they reflect the characteristics (e.g., age, gender, race/ethnicity) of people to whom generalisations are intended.

Research Limitations

Our work has several limitations that must be addressed. First, one may question the generalisability of our sample of archival cases to other criminal cases on the basis of geographical considerations. While this issue cannot be addressed conclusively at this time, it is not entirely clear why one would expect characteristics such as exposure duration or retention interval lengths to differ in a systematic way in other places or at other times. Nevertheless, we strongly urge other researchers to conduct similar archival analyses in different jurisdictions, using a greater range of crime types, to further clarify how our simulations could be improved by better reflecting a range of actual witnessing conditions. We would also welcome further archival research to compare with the present findings, especially considering the age of our sample. It would also be fruitful for researchers to work with lawyers, police, and other legal practitioners at the outset of designing their research to maximise its relevance to the legal system (Weiner et al., 2002). Second, we sampled from cases referred to the District Attorney for prosecution. The characteristics of cases that do not make it to this stage of the legal system may very well be different. Third, some may be concerned that a number of comparisons are based on coder estimations of time durations. Unfortunately, due to the nature of archival research, objective information of this sort is not always available. However, it is important to note that all coders were extensively trained and exhibited a high level of agreement when making these estimates. Fourth, there is also the question of whether missing data from the literature or from the archives may have affected our results in some way. Similarly, it is unclear that this missing information would drastically change the overall pattern of our results. We encourage researchers to provide as much detail as possible in their published reports regarding their simulations, especially for those factors which may affect accuracy either directly or indirectly. Fifth, construct validity (i.e., the certainty with which we can claim that we are measuring the theoretical concept of interest, such as, say, *witness stress*) may be compromised in archival research if the people reporting the data are not objective (e.g., the police conclude that the witness could not have been stressed, and record this in the case file, even though the witness says she was) (Wiener et al., 2002). Our observations, for instance, are based on observations made by police officers, probation officers, attorneys, and other legal officials who completed the case files. Sixth, we coded only a subset of the literature on face memory, concentrating our efforts on the studies that simulated an eyewitness memory task. Consequently, the number of studies that have investigated the effects of various phenomena (e.g., source monitoring, subject age, retention interval) on remembering is underestimated in our survey of the literature. However, had we included in our analysis the face memory papers that did not explicitly simulate an eyewitness memory task, the differences between the characteristics of the laboratory studies and the real-world identifications in all probability would have been far greater. Seventh, we did not study misdemeanour archival cases. We found that the lab studies presented mainly misdemeanour crimes, such as theft, as the to-be-remembered event. Had we compared theft in laboratory and archives, we may have found the witness ecology was similar. Finally, further work is needed to ascertain and study the types of real-world cases to which generalisations from the lab are made. We coded felony rape, robbery and assault cases because, to our knowledge, eyewitness experts frequently testify in these types of cases. Further research is needed to describe the witness ecology of other types of cases experts are called to testify about. Having said this, expert witnesses base their testimony in felony cases on all of the eyewitness literature. They do not particularise their testimony about the literature based on crime type, at least to our knowledge. Although there is no empirical evidence on the matter, in our view, crime type is probably less influential on memory accuracy than other features of crimes, such as the witness' duration of exposure to the culprit, the witness' viewing distance, and the intervals (see Ebbesen & Rienick, 1998) between when the crime occurred and the witness' testimony.

Conclusion

In the US, the view that eyewitness expert testimony is probative has rested in part on the assumption that our simulations are generalisable to actual witnessing situations. The results of this research project suggest that we need to broaden the range of conditions employed in the laboratory to increase the applicability of eyewitness identification research to the legal system in rape, robbery and assault felony cases. We have also proposed new avenues for eyewitness identification laboratory research by identifying conditions in the real-world cases that we have not yet represented in the laboratory, such as the effects of exposure to the culprit before the onset of the crime and the need for live showup research. Finally, we hope that our study will prompt additional investigations into the ecology of real-world eyewitnesses. This type of research has potential to inform both theory and applied practice.

References

Aronson, E., & Carlsmith, J. M. (1968). Experimentation in social psychology. In G. Lindzey & E. Aronson (Eds.), The handbook of social psychology (2nd ed., Vol. 2, pp. 1-79). Reading, MA: Addison-Wesley.

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

Blankertz, L. (1998). The value and practicality of deliberate sampling for heterogeneity: A critical multiplist perspective. *American Journal of Evaluation*, *19*, 307–324.

Bornstein, B. H., Deffenbacher, K. A., Penrod, S. D., & McGorty, E. K. (2012). Effects of exposure time and cognitive operations on facial identification accuracy: A meta-analysis of two variables associated with initial memory strength. *Psychology, Crime & Law, 18* (5), 473-490.

Bothwell, R. K., Brigham, J. C., & Pigott, M. A. (1987). An exploratory study of personality differences in eyewitness memory. *Journal of Social Behavior & Personality, 2* (3), 335-343.

Clifford, B. R., & Hollin, C. R. (1981). Effects of the type of incident and the number of perpetrators on eyewitness memory. *Journal of Applied Psychology, 66* (3), 364-370.

Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987). Improving the reliability of eyewitness identification: Putting context into context. *Journal of Applied Psychology, 72* (4), 629-637.

Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987). The reliability of eyewitness identification: The role of system and estimator variables. *Law and Human Behavior, 11* (3), 233-258.

Deffenbacher, K. A. (1984). Experimental psychology actually can assist triers of fact. *American Psychologist. 39*(9), 1066-1068.

Deffenbacher, K. A., Bornstein, B. H., McGorty, E. K., & Penrod, S. D. (2008). Forgetting the once-seen face: Estimating the strength of an eyewitness's memory representation. *Journal of Experimental Psychology: Applied, 14*(2), 139-150.

Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (2004). A Meta-Analytic Review of the Effects of High Stress on Eyewitness Memory. *Law and Human Behavior, 28*(6), 687-706.

De Jong, M., Wagenaar, W. A., Wolters, G., & Verstijnen, I. M. (2005). Familiar face recognition as a function of distance and illumination: A practical tool for use in the courtroom. *Psychology, Crime & Law, 11*(1), 87-97.

Dennis I (2017) *The Law of Evidence.* London: Sweet and Maxwell.

Ebbesen, E. B., & Rienick, C. B. (1998). Retention interval and eyewitness memory for events and personal identifying attributes. *Journal of Applied Psychology, 83*(5), 745-762. doi: 10.1037//0021-9010.83.5.745

Elliott, R. (1993). Expert testimony about eyewitness identification: A critique. *Law and Human Behavior, 17*(4), 423-437.

Fawcett, J. M., Russell, E. J., Peace, K. A., & Christie, J. (2013). Of guns and geese: A meta-analytic review of the 'weapon focus' literature. *Psychology, Crime & Law, 19*(1), 35-66.

Flowe, H. D., Finklea, K. M., & Ebbesen, E. B. (2009). Limitations of expert psychology testimony on eyewitness identification. In B. L. Cutler, B. L. Cutler (Eds.), *Expert testimony on the psychology of eyewitness identification* (pp. 201-221). New York, NY: Oxford University Press.

Goldstein, A. G., & Chance, J. E. (1980). Memory for faces and schema theory. The Journal of Psychology: Interdisciplinary and Applied, 105(1), 47-59. doi:10.1080/00223980.1980.9915131

Gonzalez, R., Ellsworth, P. C., & Pembroke, M. (1993). Response biases in lineups and showups. *Journal of Personality and Social Psychology, 64*(4), 525-537.

Haber, R. N., & Haber, L. (2000). Experiencing, remembering and reporting events. *Psychology, Public Policy, And Law, 6*(4), 1057-1097. doi:10.1037/1076-8971.6.4.1057

Hancock, K. J., & Rhodes, G. (2008). Contact, configural coding and the other-race effect in face recognition. *British Journal of Psychology, 99*(1), 45-56.

Hammond, K. R., & Stewart, T. R. (Eds.). (2001). *The essential Brunswik: Beginnings, explications, applications.* New York, NY, US: Oxford University Press.

Hollien, H., Bennett, G., & Gelfer, M. P. (1983). Criminal identification comparison: Aural versus visual identification resulting from a simulated crime. *Journal of Forensic Sciences*, 25, 208-221.

Horry, R., Memon, A., Wright, D. B., & Milne, R. (2012). Predictors of eyewitness identification decisions from video lineups in England: A field study. *Law and Human Behavior, 36*(4), 257-265.

Hosch, H. M., & Bothwell, R. K. (1990). Arousal, description and identification accuracy of victims and bystanders. *Journal of Social Behavior & Personality, 5*(5), 481-488

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove, CA: Wadsworth Publishing.

Kassin, S. M., Ellsworth, P. C., & Smith, V. L. (1989). The 'general acceptance' of psychological research on eyewitness testimony: A survey of the experts. *American Psychologist, 44*(8), 1089-1098.

Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the 'general acceptance' of eyewitness testimony research: A new survey of the experts. *American Psychologist, 56*(5), 405-416.

Konečni, V. J., & Ebbesen, E. B. (1986). Courtroom testimony by psychologists on eyewitness identification issues: Critical notes and reflections. *Law and Human Behavior, 10*(1-2), 117-126.

Kurosawa, K. (1996). System variables in eyewitness identification: Control experiments and photospread evaluation. *Japanese Psychological Research, 38*(1), 25-38.

Lawson, V. Z., & Dysart, J. E. (2014). The showup identification procedure: An exploration of systematic biases. *Legal and Criminological Psychology, 19*(1), 54-68.

Leippe, M. R., Wells, G. L., & Ostrom, T. M. (1978). Crime seriousness as a determinant of accuracy in eyewitness identification. *Journal of Applied Psychology, 63*(3), 345-351.

Loftus, E.F. (1983). Silence is not golden. *American Psychologist, 38*, 564-572.

Maass, A., & Köhnken, G. (1989). Eyewitness idntification: Simulating the 'weapon effect.'. *Law and Human Behavior, 13*(4), 397-408.

MacLin, O. H., MacLin, M. K., & Malpass, R. S. (2001). Race, arousal, attention, exposure and delay: An examination of factors moderating face recognition. *Psychology, Public Policy, and Law, 7*(1), 134-152.

MacLin, O. H., Van Sickler, B. R., MacLin, M. K., & Li, A. (2004). A Re-examination of the Cross-race Effect: The Role of Race, Inversion, and Basketball Trivia. *North American Journal of Psychology, 6*(2), 189-204.

Malpass, R. S., & Devine, P. G. (1980). Realism and eyewitness identification research. *Law and Human Behavior, 4*(4), 347-358.

Malpass, R. S., & Devine, P. G. (1981). Guided memory in eyewitness identification. *Journal of Applied Psychology, 66*(3), 343-350.

Malpass, R. S., Tredoux, C. G., Schreiber Compo, N. S., McQuiston-Surrett, D., Maclin, O. H., Zimmerman, L. A., & Topp, L. D. (2008). Study space analysis for policy development. *Applied Cognitive Psychology, 22*(6), 789-801.

Megreya, A. M., White, D., & Burton, A. M. (2011). The other-race effect does not rely on memory: Evidence from a matching task. *The Quarterly Journal of Experimental Psychology, 64*(8), 1473-1483.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7*(1), 3-35.

Michel, C., Caldara, R., & Rossion, B. (2006). Same-race faces are perceived more holistically than other-race faces. *Visual Cognition, 14*(1), 55-73.

Morgan, C. I., Hazlett, G., Doran, A., Garrett, S., Hoyt, G., Thomas, P., & ... Southwick, S. M. (2004). Accuracy of eyewitness memory for persons encountered during exposure to highly intense stress. *International Journal of Law and Psychiatry, 27*(3), 265-279.

Pachella, R. G. (1986). Personal values and the value of expert testimony. *Law and Human Behavior, 10*(1-2), 145-150. doi:10.1007/BF01044565

Pezdek, K. (2009). Content, form, and ethical issues concerning expert psychological testimony on eyewitness identification. In B. L. Cutler, B. L. Cutler (Eds.), *Expert testimony on the psychology of eyewitness identification* (pp. 29-50). New York, NY: Oxford University Press.

Pike, G., Brace, N., & Kyman, S. (2002). *The visual identification of suspects: Procedures and practice. Briefing note 2/02.* Policing and Reducing Crime Unit, Home Office Research Development and Statistics Directorate. Retrieved from http://webarchive.nationalarchives.gov.uk/20110220105210/http:/rds.homeoffice.gov.uk/rds/pdfs2/brf202.pdf

Riske, M. L., Wallace, B., & Allen, P. A. (2000). Imaging ability and eyewitness accuracy. *Journal of Mental Imagery, 24*(1), 137-148.

Roberts, R. (2012). Expert evidence on the reliability of eyewitness identification – Some Observations on the justifications for exclusion: *Gage v HM Advocate. International Journal of Evidence and Proof,* 16(1) 93-105.

Ryder, H., Smith, H. M. J., & Flowe, H. D. (2015). Estimator variables and memory for faces. In T. Valentine & J. P. Davis (Eds.), *Forensic facial identification: Theory and practice* *of identification from eyewitnesses, composites and CCTV* (pp.159-184). Chichester, England: John Wiley & Sons

Searcy, J., Bartlett, J. C., & Memon, A. (2000). Influence of post-event narratives, line-up conditions and individual differences on false identification of young and older eyewitnesses. *Legal and Criminological Psychology, 5*(Part2), 219-235.

Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100*(2), 139-156.

Steblay, N. M. (1992). A meta-analytic review of the weapon focus effect. *Law and Human Behavior, 16*(4), 413-424.

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. (2003). Eyewitness accuracy rates in police showup and lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 27*(5), 523-540.

Tanaka, J. W., Kiefer, M., & Bukach, C. M. (2004). A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study. *Cognition, 93*(1), B1-B9.

Tollestrup, P. A., Turtle, J. W., & Yuille, J. C. (1994). Actual victims and witnesses to robbery and fraud: An archival analysis. In D. F. Ross, J. D. Read, M. P. Toglia, D. F. Ross, J. D. Read, M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 144-160). New York, NY: Cambridge University Press.

Technical Working Group for Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement.* Washington, DC: United States Department of Justice, Office of Justice Programs.

Tredoux, C. G. (1998). Statistical inference on measures of lineup fairness. *Law and Human Behavior, 22*, 217–237.

Turtle, J. W., & Yuille, J. C. (1994). Lost but not forgotten details: Repeated eyewitness recall leads to reminiscence but not hypermnesia. *Journal of Applied Psychology, 79*(2), 260-271.

Valentine, T., Davis, J. P., Memon, A., & Roberts, A. (2012). Live showups and their influence on a subsequent video line-up. *Applied Cognitive Psychology, 26*(1), 1-23.

Valentine, T., Pickering, A., & Darling, S. (2003). Characteristics of eyewitness identification that predict the outcome of real lineups. *Applied Cognitive Psychology, 17*(8), 969-993.

Wagenaar, W. A., & van der Schrier, J. H. (1996). Face recognition as a function of distance and illumination: A practical tool for use in the courtroom. *Psychology, Crime & Law, 2*(4), 321-332.

Weisburd, D., Lum, C. M., & Petosino, A. (2001). Does research design affect study outcomes in criminal justice? *The Annals of the American Academy of Political and Social Science,578,* 50-70.

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), 603-647.

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin, 25*(9), 1115-1125.

Welsh, B. C., Peel, M. E., Farrington, D. P., Elffers, H., & Braga, A. A. (2011) Research design influence on study outcomes in crime and justice: a partial replication with public area surveillance. *Journal of Experimental Criminology, 7*(2), 183-198.

Wiener, R. L., Winter, R. J., Rogers, M., Seib, H., Rauch, S., Kadela, K., & ... Warren, L. (2002). Evaluating published research in psychology and law: A gatekeeper analysis of Law and Human Behavior. In J. P. Ogloff, J. P. Ogloff (Eds.), *Taking psychology and law into the twenty-first century* (pp. 371-405). New York, NY: Kluwer Academic/Plenum Publishers.

Wright, D. B., & McDaid, A. T. (1996). Comparing system and estimator variables using data from real line-ups. *Applied Cognitive Psychology, 10*(1), 75-84.

Yarmey, A. D. (1997). Probative v. prejudicial value of eyewitness memory research. *Expert Evidence, 5*(3), 89-97.

Yarmey, A. D. (2001). Expert testimony: Does eyewitness memory research have probative value for the courts?. *Canadian Psychology, 42*(2), 92-100.

Yarmey, A. D., Yarmey, M. J., & Yarmey, A. L. (1996). Accuracy of eyewitness identifications in showups and lineups. *Law and Human Behavior, 20*(4), 459-477.

Yuille, J. C. (1993). We must study forensic eyewitnesses to know about them. *American Psychologist, 48*(5), 572-573.

Yuille, J. C., & Cutshall, J. L. (1986). A case study of eyewitness memory of a crime. *Journal of Applied Psychology, 71* (2), 291-301.

Yuille, J. C, & Wells, G. (1991). Concerns about the application of research findings: The issue of ecological validity. In J. L. Doris (Ed.), *The suggestibility of children's recollections* (pp. 118-128). Washington, DC: American Psychological Association.

Table 1

Top 10 independent variables investigated by the laboratory studies.

	Independent Variable	Number of Experiments
1	Presence or Absence of Lineup Target	260
2	Simultaneous or Sequential Lineup	68
3	Subject Age	60
4	Lineup Instructions	47
5.5	Who target is	39
5.5	Retention Inverval	39
7	Lineup Size	34
8	What was viewed before lineup	33
9.5	Lineup Member Similarity	30
9.5	Mnemonic Techniques/Context Reinstatement	30

Table 2

Exposure duration descriptive data (in minutes) for the laboratory studies compared to the archives

Critical Event Exposure Duration (minutes)							
Studies	*assault*	*rape*	*robbery*	*theft*	*other crime*	*none*[b]	*Overall*[c]
M	0.78	1.25	1.50	1.45	1.07	3.93	2.32
SD	0.55	1.06	1.25	1.10	1.08	9.20	5.76
range	0.15 to 0.57	0.50 to 2.00	0.05 to 5.50	.05 to 5.08	.08 to 3.25	.03 to 40.0	0.03 to 40.0
median	0.57	1.25	1.03	1.25	0.75	1.00	1.00
N	9	2	64	128	23	122	335
Archives	*assault*	*rape*	*robbery*				*Overall*[c]
M	13.93	131.53	11.6				59.87
SD	51.93	611.94	39.91				390.75
range	0.08 to 540	0.50 to 7200	0.25 to 450				.08 to 7200
median	10	10	5				8
N	156	201	150				507
Target Exposure Duration (minutes)							
Studies	*assault*	*rape*[a]	*robbery*	*theft*	*other crime*	*none*	*Overall*[d]
M	0.82	1.13	0.54	0.81	0.94	3.48	1.8
SD	0.58	--	0.48	1.48	0.92	8.58	5.58
range	0.15 to 2.00	--	0.03 to 1.50	0.03 to 15.0	0.17 to 3.00	0.03 to 40.0	0.03 to 40.0
median	0.78	--	0.39	0.5	0.72	0.42	0.4
N	8	1	49	105	13	111	284
Archives	*assault*	*rape*	*robbery*				*Overall*[d]
M	9.71	133.02	16.43				65.74
SD	30.73	292.54	42.25				205.17
range	0.00 to 330	0.00 to 2423	0.23 to 231				0.00 to 242
median	2	15	5				7.8
N	146	179	78				403

a No further descriptive statistics other than those reported could be computed

b Studies in which the to-be-remembered event was not criminal

c Archive v lab: Mann-Whitney U, $z = -19.91$, $p < .0001$

d Archive v lab: Mann-Whitney U, $z = -16.41$, $p < .0001$

Table 3

Distribution of identification tasks for the witnesses in the archives and participants in the laboratory studies.

Procedure	Real World Identifications (N = 406)	Experimental Identifications (N = 93,189)[a]
Mugshots	<.01	.05
Photo Showup	.00	.02
Live Showup	.51	<.01
Photo Lineup	.40	.78
Video Lineup	.00	.09
Live Lineup	.03	.02
Voice Lineup	.00	.01
Other	.06	<.01
Unknown	.00	.02

a. Some research participants viewed mugshots and were then run in another identification procedure.

Table 4

Identification outcomes in the archives by test procedure and eyewitness/suspect level of acquaintance.

Outcome	Live Lineup n = 12	Live Showup n = 207	Photo Lineup n = 164	Other n = 23
Suspect ID	11 (7)	189 (142)	127 (85)	17 (11)
No ID made	1 (0)	8 (8)	22 (19)	2 (1)
Other	0 (0)	10 (6)	15 (9)	4 (0)

Table note: Numbers in parentheses indicate number of known stranger identifications. For instance, there were 11 live lineup suspect IDs, 7 of which were stranger IDs. "Other" identification outcomes include multiple IDs and foil IDs.

Table 5

Retention Intervals				
Studies	*description*[a]	*lineup*[b]	*showup*[c]	*any ID procedure*[d]
median	2.00 minutes	20 minutes	15 minutes	0.33 hours
M	41.79 hours	96.3 hours	45.70 hours	93.23 hours
SD	135.80 hours	317.82 hours	126.83 hours	310.7 hours
range	0 to 840 hours	0 to 3,648 hours	0 to 508.8 hours	0 to 3648 hours
Archives	*description*[a]	*lineup*[b]	*showup*[c]	
median	1.13 hours	11 days	0.96 hours	36.00 hours
M	188.08 hours	561.98 hours	223.43 hours	383.08 hours
SD	1030.58 hours	847.44 hours	1243.42 hours	1045.66 hours
range	0 to 842.40 hours	0.50 to 4992 hours	0 to 8136 hours	0 to 8136 hours

a. Studies versus archives: Mann-Whitney U, $z = -5.39$, $p < .0001$

b Studies versus archives: Mann-Whitney U, $z = -.49$, $p < .0001$

c Studies versus archives: Mann-Whitney U, $z = -3.25$, $p < .0001$

d Studies versus archives: Mann-Whitney U, $z = -2.95$, $p < .0005$