General Acceptance among Experts of Eyewitness Research Findings: Sources of Influence

Sneha Suresh
CUNY John Jay College, sneha.suresh@jjay.cuny.edu

Follow this and additional works at: https://academicworks.cuny.edu/jj_etds

Part of the Legal Studies Commons, and the Psychology Commons

Recommended Citation
https://academicworks.cuny.edu/jj_etds/73

This Thesis is brought to you for free and open access by the John Jay College of Criminal Justice at CUNY Academic Works. It has been accepted for inclusion in Student Theses by an authorized administrator of CUNY Academic Works. For more information, please contact AcademicWorks@cuny.edu.
General Acceptance among Experts of Eyewitness Research Findings: Sources of Influence

Sneha Suresh

MA Forensic Psychology Thesis

John Jay College of Criminal Justice
Abstract

Recent advances in eyewitness identification research has been rapidly expanding with emergence of new eyewitness phenomena. At the same time, the number of experts being call upon to testify in court to the reliability of these various phenomena are also increasingly. For such testimony to be admissible, it must meet various criteria of general acceptance established by the courts. In the current study, 54 experts responded to a survey to assess the extent of their consensus of general acceptance of various eyewitness phenomena, collect information on their demographics, opinions regarding various other courtroom phenomenon and assess the impact of recent literature on perceptions of reliability and general acceptance of the same phenomena. The findings revealed that there are high levels of agreement amongst experts on a variety of topics such as Showups, Blind Administration, Description Matched Lineup, Weapons Focus, Elderly Witness, High Identification Confidence, Lineup Fairness and Change Blindness. Familiarity with various eyewitness research publications was also noted to have an effect on expert’s perception of general acceptance, assessments of reliability of eyewitness phenomena and willingness to testify in court. The implications of these findings with regards to the development of eyewitness consensus and expert testimony are discussed.
Introduction

Due to the reconstructive and fallible nature of human memory, eyewitness identifications in courtroom proceedings are often inaccurate (Cutler, Dexter & Penrod, 1989; Goodman & Hahn, 1987; Penrod, Loftus, & Winkler, 1982). For instance, the National Registry of Exonerations (2016) attributes 30% of all wrongful convictions to erroneous eyewitness identifications, and the Innocence Project (2014) attributes 75% of wrongful convictions, in their exonerations, to the same. Furthermore, in State v Henderson (2011) the court stated “Study after study revealed a troubling lack of reliability in eyewitness identifications. From social science research to the review of actual police lineups, from laboratory experiments to DNA exonerations, the record proves that the possibility of mistaken identification is real. Indeed, it is now widely known that eyewitness misidentification is the leading cause of wrongful convictions across the country.” (p. 5). At the same time, eyewitness identification has also been noted as critical to the apprehension and prosecution of criminals and in the exoneration of innocent suspects (DOJ, 2017). Hence, in effort to counterbalance the pros and cons, the criminal justice system has enforced several safeguards designed to reduce false convictions resulting from mistaken identifications. Some such safeguards include the use of blind procedures, where the administrator is not involved in the investigation and does not know what the suspect looks like (e.g., Connecticut Public Act 08-143), documenting the witness’s self-reported confidence at the moment of the initial identification (Department of Justice, 2017) and jury eyewitness instructions (e.g., New Jersey v Henderson, 2011; US v Telfaire, 1972). In practice, however, the usefulness of these safeguards are limited as corroborative evidence is often lacking and effective cross-examination requires a certain degree of knowledge and understanding of human memory (Walters, 1985). Simultaneously, psychological research has demonstrated that laypersons, including police, attorneys, judges, and juries lack the required knowledge and understanding of human memory.

\footnote{Eyewitness identification procedures that, when practicable, use a double-blind administration wherein the person conducting the identification procedure is not aware of which person in the photo lineup or live lineup is suspected as being the perpetrator of the crime.}
knowledge about human memory processes to draw valid inferences about the accuracy of eyewitness identification under various witnessing conditions (Benton, Ross, Bradshaw, Thomas & Bradshaw, 2006; Brigham & Bothwell, 1983; Cutler, Penrod, & Dexter, 1987; Cutler, Penrod, & Stuve, 1988; Deffenbacher & Loftus, 1982; Noon & Hollin, 1987) and in some cases juries were found to hold beliefs, about eyewitness phenomena, that were contrary to research findings and the beliefs of experts in the field (Kassin & Barndollar, 1992). In light of the deficiencies of these safeguards, courts have turned to the use of expert psychological testimony to educate the jury about memory processes and the factors that potentially influence the memory of an eyewitness.

An expert witness must be qualified by the court before they can testify. In the U.S federal system, the qualification of an expert witnesses is controlled by Rule 702 of the Federal Rules of Evidence (Garrett & Neufeld, 2009). Under this rule, a witness may qualify as an expert witness based on knowledge, skill, experience, training or education. However, there is no explicit requirement that must be met for qualification. That is, there is no training or degree specifically required. It is at the judge’s discretion to act as a ‘gatekeeper’ and determine whether the expert is qualified to testify and whether the content of the testimony meets either the Frye or Daubert standards of expert testimony based on the information brought forward during vior dire. However, there is no precisely prescribed procedure for this process leaving the concept of ‘generally acceptable’ open to the interpretation of the judge who may have little to no knowledge on the topic (Kassin, Ellsworth, & Smith, 1989). Hence, it is likely that without appropriate guidelines, aiding in

---

2 Fed. R. Evid. 702

3 Following Daubert v. Merrell Dow Pharmaceuticals Inc. (1993) the U. S. Supreme Court granted trial judges the power to act as gatekeepers with regard to the admissibility of expert scientific evidence.

4 Frye v. United States, 293 F. 1013 (D.C. Cir. 1923). A court applying the Frye standard must determine whether or not the method by which that evidence was obtained was generally accepted by experts in the particular field in which it belongs.

5 Daubert v. Merrell Dow Pharmaceuticals, Inc., 43 F.3d 1311 (9th Cir. 1995) (1) whether the theory or technique in question can be and has been tested; (2) whether it has been subjected to peer review and publication; (3) its known or potential error rate; (4) the existence and maintenance of standards controlling its operation; and (5) whether it has attracted widespread acceptance within a relevant scientific community.
RUIPENDING HEAD: General Acceptance of Eyewitness Phenomena

outlining ‘general acceptance’ for expert testimony in court proceedings, wrongful convictions might become difficult to prevent.

In their report, Garrett and Neufeld (2009) recollect that conventional wisdom once encouraged the narrative that a sound defence and cross-examination would enable courts to properly assess the strength of forensic evidence. However, the National Academy of Science report unequivocally states, and clearly demonstrates with post-conviction DNA exonerations, that at least in criminal cases, the courts have not functioned well as ‘gatekeepers’ of questionable scientific evidence (West & Meterko, 2015), and given the lack of scientific knowledge among legal practitioners and, more importantly, judges, “judicial review, by itself, will not cure the infirmities of the forensic science community” (National Research Council, 2009, p. 12). Therefore, this burden falls on the scientific community. Moreover, it can be argued that it is unfair to expect the courts to sort through, or to assess for themselves the reliability of a device or technique outside of the legal realm, no matter how widely used. Hence, it is essential that the validity of forensic techniques be established within the respective scientific fields, before any particular piece of evidence is considered in the adjudicative process. This includes establishing and defining, within each field, including the field of eyewitness testimony, the various phenomena that are ‘generally accepted’ by experts throughout the course of time. Hence, reaching a consensus among experts about what they deem generally acceptable in the field of eyewitness testimony would aid, to a large extent, in the judge’s decision making process and also in cross-examination by attorneys. The current study aims to provide some reference to the courts regarding general acceptance of eyewitness phenomena and throw some light on what influences expert opinions of generally accepted eyewitness events.

**Eyewitness Expert Testimony and the Judiciary**

In order to truly assess the need for a consensus of eyewitness experts on various eyewitness phenomena, it is important to look at the standing of the courts on the matter. Courts are noted to
have inconsistently admitted expert testimony on the reliability of eyewitness identifications. While the overwhelming majority of courts have excluded such expert testimony in the past (e.g., *State v. Chapple*, 1983; Among federal courts, the Eleventh Circuit has been a proponent of the exclusionary rule. See *United States v. Holloway*, 1992), they are being accepted more readily in recent times (e.g., *Commonwealth of Pennsylvania v Walker*, 2014; *State v. Lawson*, 2012; *State v. Henderson*, 2011). Some frequently cited reasons for exclusion of expert eyewitness testimony include questions regarding the general acceptance and reliability of eyewitness phenomena, the argument that the information would not assist the trier of the fact, and concerns that eyewitness expert testimony could be misleading, could overwhelm the jury or lead to a battle of experts (Fradella, 2007; Stein, 2003; Woller 2003). Additionally, in their review of court rulings on the matter, Schmechel, O'Toole, Easterly & Loftus (2006) found that the most commonly cited rationale for excluding expert testimony was that it was “not beyond the ken” (p. 191) of the average juror. Furthermore, in *People v Radcliffe* (2002), Judge William C. Donnino concluded that a proponent of evidence for the admission of expert eyewitness testimony should explain whether the testimony involves “novel scientific theories and techniques” (p. 6), and if it does, it should include an offer of proof as to its general acceptance by the relevant scientific community.

In *State v. Chapple* (1983), the Arizona Supreme Court became the first in the nation to rule that the judge abused his discretion in excluding expert testimony concerning eyewitness reliability citing that the and proffered testimony was within the ken of the juror and that it would "take from the jury their own determination as to what weight or effect to give to the evidence of the eye-witness” (p. ). Similarly, in *People v McDonald* (1984), where the prosecution’s case comprised solely of eyewitness testimony, the court ruled that Dr. Robert Shomer’s testimony, regarding the processes of perception and recall, would "usurp the jury's function” (p. 363) and also that the phenomena was “not…scientific enough at this point in time” (p. 363). However, the supreme court examined and rejected the trial judge’s grounds for exclusion and concluded that standard jury
instructions would not be sufficient to convey the same sort of data that Dr. Shomer's testimony would have provided.

More recently, in *People v Lee* (2001) the defense wanted to introduce the testimony of an eyewitness expert regarding duration of an encounter, the passage of time, stress involved in an encounter, assimilation of post-incident information and race, and absence of correlation between eyewitness confidence and the accuracy of an identification. This was denied by two separate Justices, after a *Frye* hearing, and both courts maintained that corroborative evidence was sufficient for the jury to evaluate the eyewitness testimony. Upon appeal, The New York Court of Appeals ruled that the decision to admit or exclude eyewitness expert evidence is at the discretion of the trial court. The court stated that although “jurors may be familiar from their own experience with factors relevant to the reliability of eyewitness observation and identification, it cannot be said that psychological studies regarding the accuracy of an identification are within the ken of the typical juror” (p. 10). This ruling is more in line with research. Benton, Ross, Bradshaw, Thomas & Bradshaw (2006) examined a sample of jurors, judges and law enforcement professionals on their knowledge of factors affecting eyewitness accuracy. They found participant responses differed significantly from responses of eyewitness experts. Jurors disagreed with the experts on 87% of the issues, while judges and law enforcement disagreed with the experts on 60% of the issues. The findings showed a large deficiency in knowledge of eyewitness memory amongst jurors, judges and law enforcement personnel, indicating that the legal system may benefit from expert assistance in the evaluation of eyewitness evidence.

Unlike the typical rulings on the matter, in *People v Smith* (2002), Judge James A. Yates denied the prosecution’s request for a *Frye* hearing to determine if the proffered expert eyewitness testimony, concerning “stress, event violence, cross-racial impact, and exposure time that affect the reliability and accuracy of eyewitness identification,” (p. 21) and ruled in favor of expert testimony stating that it was generally accepted in the relevant scientific community. In this case, the judge’s
decisions was criticized by Fried (2018) for having relied solely on articles which had urged such
testimony, and not addressing the numerous studies, “whose methodologies, measurements and
conclusions were contested amongst the experts in the relevant scientific community” (p. 22) and
implicitly accepting its general acceptability. He further stated that while the New York court
decisions relied on by Smith included cases which had admitted similar testimony as beyond the ken
of the jury or for other reasons, perhaps implicitly finding general acceptability, there were no
explicit findings concerning general acceptability. Moreover, the previous New York cases which
had declined to admit such testimony generally did so on the ground that such evidence was within
the ken, thereby failing to make any findings on general acceptability.

Other key rulings include People v. LeGrand (2007) were the defense wanted to introduce
the testimony of an eyewitness expert regarding the effect of weapon focus, the lack of correlation
between witness confidence and accuracy of identification, the effect of post-event information on
accuracy and confidence malleability. The prosecution opposed the motion on the ground that the
theories that the defense wanted to introduce were "highly controversial” (p. 8). In State v
Henderson (2011) the court appointed a Special Master to evaluate scientific and other evidence
about eyewitness identifications. These factors included system variables like lineup procedures and
estimator variables like lighting conditions or the presence of a weapon. In this case, the court
concluded that the existing standard for assessing eyewitness identification evidence did not offer
an adequate measure for reliability, did not sufficiently discourage inappropriate police conduct and
that it also overstates the inherent ability of the jury to evaluate the testimony of eyewitnesses who
honestly believe in its accuracy. Having reviewed the Special Master’s report, the New Jersey
Supreme Court determined that if the defendant can show evidence of suggestibility, all system and
estimator variables relevant to the identification should be investigated at a pretrial hearing, and if
the prosecution demonstrates that the identification is still reliable, the identification should be
admitted. Furthermore, the court stated that upon admittance of the identification, the jury must be
instructed on the research concerning the relevant system and estimator variables in order to help
the jurors evaluate the reliability of the identification. Furthermore, the court stated “We anticipate,
however, that with enhanced jury instructions, there will be less need for expert testimony.” (p.
126).

Interestingly, research into such jury instructions, including Henderson, suggest that they are
not very effective in aiding the jury (Papailiou, Yokum & Robertson, 2015; Cutler, Dexter &
Penrod, 1989; Cutler, Dexter & Penrod, 1990; Greene, 1988; Jones & Penrod, 2018; Jones, Bergold,
Dillon & Penrod, 2017; Dillon, Jones, Bergold, Hui & Penrod, 2017; Jones, Bergold, Dillon &
Penrod, 2017). On the other hand, research evaluating the effectiveness of eyewitness expert
testimony has been more optimistic (Devenport, Stinson, Cutler & Kravitz, 2002; Cutler, Penrod &
now sound reasons to believe that jurors not only need [expert] testimony [on eyewitness
identification] but [that they] also benefit from it” (p. 29) and Loftus & Doyle (1997) “Recent
research ... suggests that while expert testimony is no panacea, it does enhance the quality of jury
deliberations” (p. 29) to shed light on the importance of expert eyewitness in the courtroom.

In Oregon v. Lawson (2012) the Oregon Supreme Court ruled that in situations in which the
defendant objects to the veracity of the the eyewitness testimony, the burden is on the prosecution to
show its validity. In Commonwealth of Pennsylvania v Walker (2014), the Appellant’s motion to
admit expert testimony on human recall, or to hold a Frye hearing was denied. The trial court relied
upon Pennsylvania case law which held that expert testimony concerning eyewitness identification
was inadmissible at the time. Upon appeal, the Supreme Court ruled “we hold that, in
Pennsylvania, the admission of expert testimony regarding eyewitness identification is no longer
per se impermissible, and join the vast majority of jurisdictions which leave the admissibility of
such expert testimony to the discretion of the trial court.” (p. 1). Thus, reversing the order of the
Superior Court, and remanding the matter to the trial court for reconsideration of such expert testimony in the future including the possibility of a *Frye* hearing.

Given the court’s long-standing apprehension regarding what lies within the ‘ken of the juror’ as well as what can be deemed scientifically accepted at the time, it would be helpful to have an appraisal of experts on the matter. Specifically, considering jurisdictions under Frey and Daubert standards, and expert consensus on ‘generally accepted’ eyewitness phenomena and an assessment of the relationship between perceptions of reliability and quality of research available that influence the formation of scientifically based opinions could surely aid the courts in their decisions.

**Expert Eyewitness Consensus**

The earliest effort to address the question of expert consensus of general acceptance was spearheaded by Yarmey and Jones (1983). They had 16 eyewitness experts read hypothetical situations that were designed to assess their opinions on a variety of topics. The experts were then asked to choose one out of four statements that best described the likely outcome. The results of this study revealed high levels of consensus on many topics. Unfortunately, as pointed out by Konecni & Ebbesen (1986), Yarmey and Jones (1983) failed to address the perceived reliability of the findings as the experts were not provided with an alternative response that enabled them to characterize the effects as weak, unreliable, or mixed.

Kassin, Ellsworth and Smith (1989) conducted a more focused research attempting to explore and document expert consensus on eyewitness phenomena that are ‘generally acceptable’. They compiled a list of experts by searching the various eyewitness literature published in scientific journals between 1980 and mid-1986 and surveyed them regarding their opinions on a list of 21 eyewitness related topics. Their research established that there does indeed exist a consensus amongst eyewitness experts about ‘acceptable’ phenomena. They found a strong consensus of 80% on the research data on various phenomena (wording of questions, lineup instructions, misleading post event information, the accuracy-confidence correlation, attitudes and expectations, exposure
time, unconscious transference, showups, and the forgetting curve) were reliable enough to be presented in court. Furthermore, over 70% of the experts also endorsed lineup fairness, the tendency to overestimate the duration of events and the cross-race identification bias among White witnesses. Most eyewitness experts who had testified reported to have done so on behalf of criminal defendants. They also reported they would be just as likely to consent for the prosecution as for the defense. Moreover, they reported to be more likely to agree to testify in civil cases than in criminal. Concerning their role in court, most respondents indicated that their main objective is to educate the jury, and that juries are more competent with the aid of experts than without.

The former study was updated with a very similar latter study by Kassin, Tubb, Hosch and Memon (2001) where they surveyed sixty-four psychologists about their courtroom experiences and opinions on 30 eyewitness phenomena. In light of Daubert ruling in 1993, they also asked respondents to indicate for each item whether their opinion was based on published, peer-reviewed, scientific research and found 80% consensus among experts about sufficiently reliable eyewitness phenomena that could be presented in court. These phenomena involved the wording of questions, lineup instructions, confidence malleability, mug-shot-induced bias, post event information, child witness suggestibility, attitudes and expectations, hypnotic suggestibility, alcoholic intoxication, the cross-race bias, weapon focus, the accuracy-confidence correlation, the forgetting curve, exposure time, presentation format, and unconscious transference. Additionally, the results also indicated that these experts set high standards before agreeing to testify.

The findings form Kassin, Tubb, Hosch and Memon (2001) survey has been acknowledged by the courts (e.g., New Jersey v. Henderson, 2011; People v Legrand, 2002; People v Williams, 2006). Specifically, People v Legrand (2002) and People v Williams (2006) acknowledge its usefulness in the testimony of Dr. Malpass regarding the relationship between confidence and accuracy since its publication. However, the main issue cited by both these courts regarding the
survey, and also acknowledged by the authors, was its low response rate of 34% which is a weak representation of eyewitness experts.

Daftary and Penrod (2012) conducted a similar study in two waves; the first in 2007 and the second in 2012 and reports a useable response rate of 79% for the first wave and 55% for the second wave. A comparatively stronger representation of eyewitness experts. A primary goal of Daftary and Penrod (2012), for both waves, was to assess expert’s familiarity with a number of research studies that included primary research and meta-analyses, on various eyewitness phenomena, and the impact of these research articles on their opinions of general acceptance and willingness to testify. A secondary goal was to provide an update to Kassin, et al. (2001). Both goals focused on topic areas in which substantial research had been completed since 2001 and where there might be a detectable shift in opinions amongst experts in the field. Consistent with previous research, they found a high consensus among experts regarding reliability of eyewitness Phenomena. The study also found that the level of expertise of experts, determined by the number of publications and their testifying experience, were more willing to testify to the various phenomena in court. Additionally, experts who rated the research in a particular area as being of high quality and generally reliable were more willing to deem the phenomena as generally acceptable in the field and to testify in court. Strong correlations between ratings of general acceptability (“generally reliable” or “very reliable”) and willingness to testify were also found. Specifically, of those experts who rated a phenomenon as very reliable 97% were willing to testify to it in court, and for those who rated a phenomenon to be generally reliable 89% were willing to testify to the phenomenon in court. Interestingly, experts who were more familiar with meta-analyses regarding various phenomena were noted to rate those phenomena as more reliable, as compared to those who were unfamiliar with the phenomena in question (e.g. mugshot induced bias, unconscious transference, presentation format, and stress). Notably the results suggest a substantial increase in requests to testify as well as a decline in acceptance rates. Hence, Daftary
and Penrod (2012) suggest the possibility that experts are receiving more requests than they can handle. Additionally, experts cited lack of time, perceived level of incompetence on the particular area, or moral and ethical reasons for declining to testify.

The current study aims to update Daftary and Penrod (2012) for the third wave by incorporating new phenomena and published research articles that have not been surveyed before. This wave will focus on meta-analyses and most-cited recent articles and their influence on expert opinion formation. This will provide a reference for courts regarding research phenomena that have come about in the recent past and also provide scholars with insight into the familiarity and influence of different types of research articles.

The Current Study

In a recent opinion of the Supreme Court of New Jersey, the Supreme Court of New Jersey went noted that “in the ten years since the Kassin study, the consensus that the study of eyewitness identification is a reliable field of research has continued to grow.” (p. 30; Tillman v State, 2011). Indeed, the field of eyewitness research has arguably made significant advances in recent years since the initial surveys (Kassin, Ellsworth & Smith, 1989; Kassin, Tubb, Hosch & Memon, 2001; Daftary & Penrod, 2012) of eyewitness experts. There have been a number of new eyewitness phenomena that have come to light, new published research articles and meta-analyses which could potentially impact experts’ perceptions of various eyewitness phenomena since the previous survey. Additionally, phenomena previously surveyed that yielded low consensus where new publications are available could also benefit from an update on expert consensus. These include research on description matched lineup (Fitzgerald, Price, Oriet & Charman, 2013), elderly witnesses (Erickson, Lampinen & Moore, 2016), lineup fairness (Fitzgerald, Price, Oriet & Charman, 2013; Fitzgerald, Oriet & Price, 2015), presentation format (Palmer & Brewer, 2012), showups (Neuschatz, Wetmore, Key, Cash, Gronlund &Goodsell, 2016; Wetmore, Neuschatz, Gronlund, Wooten, Goodsell & Carlson, 2015), weapons focus (Fawcett, 2013; Kocab & Sporer, 2016), blind
administration (Charman & Quiroz, 2016), change blindness (Fitzgerald, Oriet & Price, 2016), own gender bias (Rhodes & Anastasi, 2012) and high identification confidence (Wixted & Wells, 2017). An updated survey of this nature can greatly assist judges in determining the quality of available eyewitness research, as well as its general acceptance in the scientific community. Furthermore, in light of findings from Daftary & Penrod (2012) further exploring expert’s familiarity of meta-analysis and its influence on their opinion formation could allow for a better understanding of expert’s appraisal of available research and possibly aid in the production of better research material. Furthermore, this could also throw some light on the emergence of consensus, expanding our knowledge and understanding of the same.

**Methods**

**Participants**

The survey involved the initial list of experts with locatable emails compiled for the first two waves by Daftary and Penrod (2012). This included a list of 100 experts known to the authors of Daftary and Penrod (2012) based on prior published research compiled and was supplemented with names generated in a search of PsycINFO for eyewitness-related research published from 2001 to June 2007 (n=399 articles) for the first wave, and 20 additional authors, who met the same selection criteria, for the second wave. For the third wave recent research articles were culled for additional names (of senior authors where it was possible to make this determination) which brought the total number of contacts to 175, 19 of which were returned as undeliverable—13 more recent email addresses were located among that group for a total of 169 solicitations.

Respondents were sent an email requesting their participation with a link to the survey. The survey was hosted on a secure website (www.qualtrics.com), and all responses were kept confidential.

**Materials**
A survey designed after the Daftary and Penrod (2012) survey instrument was used. The survey contained two questions regarding their participation in the 2007 survey (yes/no) and the 2012 survey (yes/no), 11 statements concerning eyewitness phenomena (see table 1), respondent ratings for perceived influence of scientific research on jury verdicts for each phenomenon, 11 demographic questions (involving qualifications, areas of expertise, number of articles published and courtroom experience) and 11 questions on article familiarity and influence on expert decision making (“how familiar are you with this article listed below?” and “to what extent did the above mentioned study influence your perception of the reliability of the relationships evaluated above (note that if, for example, you were already quite convinced about the reliability of a phenomenon, a new, confirming study might have little effect”)). For the 11 statements concerning eyewitness phenomena, respondents were asked to rate their beliefs regarding the reliability of a number of statements related to the 11 eyewitness phenomena on a 7 point Likert scale (based on Kassin, et al. 1989) that included (1) “evidence suggests the reverse is possibly true”, (2) “evidence does not support it”, (3) “evidence is inconclusive”, (4) “evidence tends to favour it”, (5) “evidence is generally reliable”, (6) “evidence is very reliable” and (7) “I don’t know”. Respondents were asked to indicate whether they thought the phenomenon was reliable enough to present in courtroom testimony (yes/no), whether they would be willing to testify to the reliability of the phenomenon (yes/no), whether their opinions were based on peer reviewed published research (yes/no), and finally they were asked for their opinions regarding whether “most jurors believed the statement to be true as a matter of common sense.” (yes/no). Experts were also asked to rate, on a 9 point Likert scale ranging from not at all to extensively, to what extent they think jury verdicts are influenced by a serious consideration of each of the 11 phenomena. Expert opinion on an appropriate effect size to reflect reliable findings in a study were also enquired for Pearson’s r (r ranging from -1 to 1) and for Cohen’s d (d ranging from 0 to 2).
Experts were asked additional questions regarding their courtroom experience: the estimated number of times they have been asked to testify, the number of times they agreed to testify, and the number of times they actually testified. Reasons for not testifying based on responses from Daftary and Penrod (2012) were also included in the survey (I'm not interested, I don't have time / it is too time consuming, it is too much of a hassle, I don't feel I have sufficient depth or breadth of knowledge in the field, the science is not strong enough in this field, I do not testify regardless of

<table>
<thead>
<tr>
<th>Phenomena</th>
<th>Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retained from Daftary and Penrod (2012)</td>
<td></td>
</tr>
<tr>
<td>Description Matched Lineup</td>
<td>The more the members of a lineup resemble a witness's description of the culprit, the more accurate an identification of the suspect is likely to be</td>
</tr>
<tr>
<td>Elderly Witness</td>
<td>Elderly witnesses are less accurate than younger adults</td>
</tr>
<tr>
<td>Lineup Fairness</td>
<td>Greater suspect similarity in a lineup tends to produce more more discriminating eyewitness judgments</td>
</tr>
<tr>
<td>Presentation Format</td>
<td>Sequential lineups produce more discriminating identifications when compared to simultaneous lineups</td>
</tr>
<tr>
<td>Showups</td>
<td>The use of a one-person showup instead of a full lineup increases the risk of misidentification</td>
</tr>
<tr>
<td>Weapons Focus</td>
<td>The presence of a weapon impairs an eyewitness's ability to accurately identify the perpetrator’s face</td>
</tr>
<tr>
<td>Blind Administration</td>
<td>Procedures wherein the lineup administrators does not know the identity of the suspect reduces false identifications</td>
</tr>
<tr>
<td>Change Blindness</td>
<td>Change blindness can lead to false identification of innocent bystanders</td>
</tr>
<tr>
<td>2017 Survey Only</td>
<td></td>
</tr>
<tr>
<td>Own Gender Bias</td>
<td>Witnesses are more accurate with same gender identification</td>
</tr>
<tr>
<td>Own Age Bias</td>
<td>Witnesses are more accurate in recognising suspects within their own age range</td>
</tr>
<tr>
<td>High Identification Confidence</td>
<td>There is a strong positive correlation between confidence and accuracy</td>
</tr>
</tbody>
</table>
the case, I am involved only in research, I have never been asked to testify, I do not approve of specific court procedures in this area and other (please specify)).

Participants were asked a number of professional demographic questions and questions concerning their background. They were also asked to indicate the highest graduate degree they had obtained, their primary area of specialization, and professional memberships. Additionally, they were asked to indicate scholarly achievements related to the area of eyewitness research, such as peer-reviewed journal articles, law review articles, books, book chapters, magazine and newsletter articles and other.

Respondents were asked to specify how often they had been called in criminal and civil cases and by which side (i.e. the prosecution or the defense). They were also asked a few general follow-up questions patterned on Kassin et al. (2001) and Daftary and Penrod (2012): (1) “What do you see as the primary role of the eyewitness expert: to educate the jury, assist a particular party, or other (please specify)?” and (2) “In general, would you say that juries are better equipped to evaluate eyewitness testimony with or without the aid of a competent expert (or is there no difference)?”

Finally, respondents were asked to assess the impact of selected research articles published between 2012 and 2017, on their judgments of reliability of research publications. Publications chosen represented some of the recent research with preference for meta-analysis, where available, and larger number of citations. For each article, experts were asked to indicate their familiarity and level of influence on their opinion of reliability (To what extent did the above-mentioned study influence your perception of the reliability of the relationships evaluated above (note that if, for example, you were already quite convinced about the reliability of a phenomenon, a new, confirming study might have little effect).

Procedure
In order to produce data on expert consensus and assemble an acceptable list of phenomena for the survey, some phenomena, where expert consensus varied in the past, were retained (repeat phenomena; change blindness, elderly witness, lineup fairness, presentation format and showups) from Daftary and Penrod (2012) survey and new phenomena, that had not been surveyed before were added (own gender bias, own age bias and high identification confidence). The process for selection and elimination of phenomena from Daftary and Penrod (2012) was based on (1) retaining phenomena that had varied responses for reliability ratings and eliminating phenomena that had mostly high or mostly low reliability ratings (2) prioritizing retaining phenomena where a new-meta analysis has been published since 2012 and consensus of general acceptance was not previously established. This was done to avoid a repeated survey of phenomena where general acceptance had already been established and was unlikely to change (as no contradicting research of significance was found to have surfaced) and assuming that the emergence of a new meta-analysis would aid in establishing consensus where consensus had previously varied.

The new eyewitness topics selected for the study were chosen from areas of eyewitness research in which there had arguably been significant new research reported since Daftary and Penrod (2012), and where there might be a detectable shift in opinions amongst experts in the field based on the emergence of some contradicting research in areas where notably high consensus is known to have not been achieved among experts (change blindness and its effects on eyewitness testimony for instance). However, due to the emergence of a number of new eyewitness phenomena since 2012 and the inability to survey them all, a more restrictive approach was adapted. Phenomena were chosen with preference for those that had a recent meta-analysis published. A PsychInfo search with the phrase “eyewitness meta-analysis” was done to obtain a list of meta-analyses published between 2012 and 2017. Phenomena with a meta-analysis were given preference over other phenomena for several reasons (1) the findings from Daftary and Penrod (2012) suggest meta-analysis were more influential when compared to primary research articles (2) a substantial
amount of research and interest in the area would be required for the emergence of a meta-analysis in the particular area of research (3) meta-analysis could be considered a possibly influential source of research in the development of expert opinions as it often condenses the findings of several primary research studies to produce an arguably more-rounded conclusion. Overall, 11 phenomena were assembled (8 retained from Daftary and Penrod (2012) and 3 new phenomena).

A PsychInfo search was then performed for each phenomenon (restricting the search to articles published since 2012) which produced a list of several articles. Meta-analysis, where available, were selected for each phenomenon with preference given to articles with high citation counts. Articles that were cited more often were retained and articles that were cited less often were eliminated. Citation count was used in the selection process to ensure a fair comparison between the expert’s familiarity with the article and its influence over their reliability ratings of the 11 phenomena tested in the survey (it was reasoned it would be more likely that respondents would recognize articles that were cited more often than those that were cited less often).

The respondents were sent an email requesting their participation with a link to the survey. The survey was not randomized as the statements introducing the phenomena and its corresponding ratings had to be quarried before the articles were presented to ensure that the respondents were familiar with the statements for each phenomenon. A brief explanation about the study and its potential uses were provided and consent was obtained through a ‘yes, I consent’ and ‘no, I do not consent’ options. The survey was anonymous and required approximately of 28 minutes to complete. A total of 54 usable responses were received, yielding a response rate of 31.95%.

**Results**

**The Experts**

Of the 54 experts who responded to the questions regarding participation in the previous two surveys and also rated various questions regarding the 11 eyewitness phenomena, 2 had only participated in 2007 survey (Wave I of Daftary & Penrod, 2012), 11 had only participated in the
2012 survey (Wave II of Daftary, 2012), 8 participated in both surveys (Wave I and Wave II of Daftary & Penrod, 2012) and 28 in neither of the previous two surveys.

Among the 44 experts who completed the demographic section, 43 held Ph.D.’s and 1 held D.Sc. The respondents listed their primary areas of specialisation as eyewitness identification (34%), memory (20%), cognitive psychology (16%), eyewitness memory (11%), social psychology (9%), eyewitness testimony (6%), and miscellaneous other areas (jury decision making, psychology and law, investigative psychology, child and elderly witness). Furthermore, experts reported having professional memberships with APLS (43%), APA (29%), SARMAC (17%), APS (11%) and other organisations (CPA, Psychomonics, BPS, EAPL and ANZAPPL). On average, the 45 experts who responded to the questions regarding their scholarly achievements had published 25.73 scientific journal articles (n=1,132; 13.22 in Kassin, et al., 2001 and 30.71 in Daftary & Penrod, 2012), 3.70 Law review articles (n=163; 1.42 in Kassin, et al., 2001 and 4.89 in Daftary & Penrod, 2012), 1.36 books (n=60; 2.15 in Kassin, et al., 2001 and 3.21 books in Daftary & Penrod, 2012), 8.64 book chapters (n=380; 6.54 in Kassin, et al., 2001 and 9.97 in Daftary & Penrod, 2012), 1.27 magazine articles (n=56; 5.38 in Kassin, et al., 2001), 0.64 newsletter articles (n=28; 5.38 in Kassin, et al., 2001) and 1.30 other publications (n=57). Experts seem to have produced more publications in Daftary & Penrod (2012; M=12.19) than in the current survey (M=9.85), however this difference was not statistically significant, t(6)=.084, p=.790. Respondents also reported reading an average of 30 eyewitness related articles and attend an average of 2 conferences a year.

In addition, many of the respondents reported being actively involved in the judicial system. Eighty two percent had been asked to testify as eyewitness experts on at least one occasion, an estimated average of 98.84 times. Overall, the sum of reported numbers was 4,133 (at least; with a maximum report of 1000+ and minimum of 0 times; n=1 and 8 respectively). Daftary and Penrod (2012) reported an overall 2,719 at Wave I and 4532 at wave II. In cases where experts were asked to testify, they agreed an estimated 77% (78% in Kassin, et al., 2001 and 66% in Daftary & Penrod,
2012) of the time and actually testified 64% of the time (92% in Kassin, et al., 2001 and 88% in Daftary & Penrod, 2012). Overall, experts reported involvement in 91% criminal cases (n=2,512; 93% 74% for defense only, 4% for prosecution only and 22% reported involvement for both sides on different occasions) and 9% civil cases (n=262; 69% for plaintiff only, 15% for defendant only and for both sides). Kassin et al. (2001) reported 93% involvement in criminal and 7% involvement in civil cases. Daftary and Penrod (2012) reported 96.8% involvement in criminal and 46% in civil cases. Both Kassin et al. (2001) and Daftary and Penrod (2012) also reported similar disproportionate requests by the defense and prosecution in criminal proceedings. Furthermore, expert testimony was admitted in court an average of 46% of the time when the attorney actively sought testimony. Reasons given by respondents for not testifying included “I'm not interested” (8%), “I don't have time / It is too time consuming” (49%), “It is too much of a hassle” (18%), “I don't feel I have sufficient depth or breadth of knowledge in the field” (18%), “The science is not strong enough in this field” (13%), “I do not testify regardless of the case, I am involved only in research” (5%), “I have never been asked to testify” (10%), “I do not approve of specific court procedures in this area” (3%) and others notable reasons involved the expert testimony being unnecessary or not beneficial to the case and scheduling conflicts (5% in both cases). For the first 11 responses, the options did not allow multiple response and hence respondents were unable to check more than one reason for not testifying. In some cases, the respondents provided the additional reason(s) in the text response box provided for “other reasons” in the survey. These responses were made note of while calculating the overall percentages for each option.

Ninety three percent (n=40; 77% and n=49 in Kassin, et al., 2001) of the respondents reported that the primary role of the expert was to educate the jury and 7% (n=3; 5% and n=3 in Kassin, et al., 2001) reported that it was to assist a particular party. Additionally, experts whose country did not have a jury system reported that their main role was to educate the judge or the court. Sixty four
percent of experts believe that expert testimony is helpful to the jury while evaluating eyewitness testimony. Moreover, experts were significantly more likely to be willing to testify in court when they regarded a phenomena as reliable enough to be presented in a courtroom, and less likely to be willing to do the same when they reported the phenomena was not reliable enough to be presented in a courtroom; \( r(44) = .903, p < .001 \). Interestingly, their willingness to testify (“Would you be willing to testify in court that this phenomenon is reliable?”; \( M=12.20, SD=4.50 \)) and likeliness to regard a phenomena as reliable enough (“Do you believe the phenomenon is reliable enough for psychologists to present it in courtroom testimony?”; responses summed across phenomena; \( M=12.20, SD=4.49 \)) was not influenced by their overall courtroom experience (sum of number of times they were asked to testify, agreed to testify and actually testified; \( M=201.61, SD=525.35 \)); \( r(44) = .04, p < .776 \) and \( r(44) = .065, p < .676 \) respectively. Similarly, expert’s opinions regarding phenomena being reliable enough for psychologists to present in court was also not influenced by their research productivity (sum of all reported publications; \( M=42.64, SD=53.13 \)) ; \( r(44) = -.286, p < .060 \). However, expert’s willingness to testify was significantly negatively correlated to their research productivity, \( r(44) = .342, p < .023 \). Experts with higher research productivity were more willing to testify than those with lower research productivity. Hence, these results suggest that expert opinions on reliability of phenomena is independent of their courtroom experience and research productivity. Expert’s willingness to testify on the other hand, while independent of courtroom experience, is influenced by their research productivity. The individual phenomena ratings for reliability and willingness to testify are reported in table 2.

Self assessed expertise (research productivity and courtroom experience combined) as a whole, yielded non-significant influence over reliability and willingness to testify; \( r(44)=.015, p<.924 \) and \( r(44)=.029, p<.851 \) respectively. These results were dissimilar with those revealed by Daftary and Penrod (2012) at Wave I where respondents with more expertise were found to be more
likely to endorse general acceptance of eyewitness phenomena ($\beta = .28$, $S.E. = 1.50$, 95% C.I. [.15, .40], $p < .01$), and more willing to testify ($\beta = .14$, $S.E. = .20$, 95% C.I. [.12, .15], $p < .05$).

**Judgments of Eyewitness Phenomena**

The reliability ratings ranging from the 11 phenomena were compiled and their acceptance ratings are reported in table 3 with the corresponding ratings from the two waves of Daftary and Penrod (2012). Overall, 54 experts responded to this section of the survey. Consistent with the findings of Kassin et al. (2001) and Daftary and Penrod (2012), the results suggest that there does exist a general consensus among experts regarding eyewitness phenomena.

Experts were asked two questions regarding their beliefs about what is within the *ken* of the jury: (1) they were asked to state their opinions regarding the extent to which they thought jury beliefs about each of the 11 phenomena statements to be true as a matter of common sense on a
Table 3. Distribution of reliability judgements for eyewitness phenomena queried.

<table>
<thead>
<tr>
<th>Phenomena</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description Matched Lineup</td>
<td>-/0/2</td>
<td>-/14/4</td>
<td>-/7/8</td>
<td>-/8/11</td>
<td>-/47/43</td>
<td>-/21/26</td>
<td>-/3/5</td>
</tr>
<tr>
<td>Elderly Witness</td>
<td>0/0/0</td>
<td>0/5/6</td>
<td>17/15/10</td>
<td>29/22/26</td>
<td>31/32/36</td>
<td>15/18/16</td>
<td>8/7/6</td>
</tr>
<tr>
<td>Lineup Fairness</td>
<td>8/8/11</td>
<td>5/17/13</td>
<td>12/12/16</td>
<td>14/8/11</td>
<td>33/29/29</td>
<td>14/13/9</td>
<td>14/12/11</td>
</tr>
<tr>
<td>Presentation Format</td>
<td>0/0/4</td>
<td>0/0/16</td>
<td>10/13/27</td>
<td>20/23/24</td>
<td>42/28/13</td>
<td>26/36/16</td>
<td>2/2/0</td>
</tr>
<tr>
<td>Showups</td>
<td>0/0/0</td>
<td>3/2/0</td>
<td>10/5/0</td>
<td>25/18/9</td>
<td>28/38/30</td>
<td>29/32/62</td>
<td>5/3/0</td>
</tr>
<tr>
<td>Weapons Focus</td>
<td>0/-/0</td>
<td>2/-/0</td>
<td>3/-/11</td>
<td>28/-/20</td>
<td>34/-/40</td>
<td>31/-/30</td>
<td>2/-/0</td>
</tr>
<tr>
<td>Blind Administration</td>
<td>-/0/0</td>
<td>-/0/0</td>
<td>-/3/0</td>
<td>-/19/11</td>
<td>-/26/30</td>
<td>-/52/60</td>
<td>-/0/0</td>
</tr>
<tr>
<td>Change Blindness</td>
<td>-/0/0</td>
<td>-/2/0</td>
<td>-/16/13</td>
<td>-/16/35</td>
<td>-/19/26</td>
<td>-/10/11</td>
<td>-/37/15</td>
</tr>
<tr>
<td>Own Gender Bias</td>
<td>-/-/0</td>
<td>-/-/23</td>
<td>-/-/26</td>
<td>-/-/21</td>
<td>-/-/4</td>
<td>-/-/0</td>
<td>-/-/26</td>
</tr>
<tr>
<td>Own Age Bias</td>
<td>-/-/0</td>
<td>-/-/2</td>
<td>-/-/9</td>
<td>-/-/47</td>
<td>-/-/17</td>
<td>-/-/9</td>
<td>-/-/17</td>
</tr>
<tr>
<td>High Identification Bias</td>
<td>-/-/4</td>
<td>-/-/24</td>
<td>-/-/18</td>
<td>-/-/18</td>
<td>-/-/18</td>
<td>-/-/18</td>
<td>-/-/0</td>
</tr>
</tbody>
</table>

Note. Table reports percentage of experts whose opinion reflects 1=Evidence suggests reverse is possibly true, 2=Evidence does not support it, 3=Evidence is inconclusive, 4=Evidence tends to favor it, 5=Evidence is generally reliable, 6=Evidence is very reliable and 7=I don’t know.
scale ranging from 0-“Jurors do not believe the above statement to be true” to 100-“Jurors believe the above statement to be true” and (2) they were asked to indicate to what extent they thought jury verdicts were influenced by a serious consideration of each of the 11 eyewitness phenomena on a scale from 1 (not at all) to 9 (extensively). The mean distribution of (1) and (2) for each phenomenon are reported in table 4.

Table 4. Mean expert ratings for perceived ‘ken’ of jury.

<table>
<thead>
<tr>
<th>Phenomena</th>
<th>True as a matter of common sense?*</th>
<th>Jury verdicts influenced by a serious consideration?**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description Matched Lineup</td>
<td>47.66</td>
<td>2.80</td>
</tr>
<tr>
<td>Elderly Witness</td>
<td>63.27</td>
<td>5.07</td>
</tr>
<tr>
<td>Lineup Fairness</td>
<td>47.95</td>
<td>4.44</td>
</tr>
<tr>
<td>Presentation Format</td>
<td>27.08</td>
<td>2.75</td>
</tr>
<tr>
<td>Showups</td>
<td>39.68</td>
<td>3.64</td>
</tr>
<tr>
<td>Weapons Focus</td>
<td>41.79</td>
<td>4.29</td>
</tr>
<tr>
<td>Blind Administration</td>
<td>38.82</td>
<td>3.22</td>
</tr>
<tr>
<td>Change Blindness</td>
<td>24.95</td>
<td>2.29</td>
</tr>
<tr>
<td>Own Gender Bias</td>
<td>35.22</td>
<td>2.60</td>
</tr>
<tr>
<td>Own Age Bias</td>
<td>40.46</td>
<td>3.00</td>
</tr>
<tr>
<td>High Identification Confidence</td>
<td>71.59</td>
<td>8.91</td>
</tr>
</tbody>
</table>

* rated on scale ranging from 0-“Jurors do not believe the above statement to be true” to 100-“Jurors believe the above statement to be true”
** rated on a scale ranging from 1-“not at all” to 9-“extensively”

Experts were also noted to have based their opinions of phenomena reliability (“Do you believe the phenomenon is reliable enough for psychologists to present it in courtroom testimony?”), willingness to testify (Under the right circumstances, would you be willing to testify in court that this phenomenon is reliable?) and general acceptance of phenomena on their scientifically based opinions (“Is your opinion in this matter based on published, peer reviewed, scientific research?”); rs (486)=.402, p < .001, rs (486)=.400, p < .001 and rs (486)= .109, p < .017.
That is, they were more likely to deem a phenomena reliable enough and testify regarding the phenomena in court when they believed their opinions were based on scientific research and likely to find the phenomena not reliable enough and be unwilling to testify when they believed their opinions were not based on scientific research. A similar relationship was found between general acceptance of phenomena with expert’s willingness to testify; \( r_s = (487) = -.385, p < .001 \).

**Judgements of Appropriate Effect Size for Reliable Research Findings**

Respondents were asked to indicate their expected appropriate effect size that reflects reliable findings in a study for Pearson’s \( r \) and for Cohen’s \( d \) (“What would you expect to be a appropriate effect size to reflect reliable findings in a study?”). Overall, for Pearson’s \( r \) expert’s reported 0.30 and 0.40 (\( n=7, 17.4\% \) in both cases) to be the most appropriate effect size. Expert judgements of effect size for Cohen’s \( d \) was more varied with the most ratings for 0.50 and 0.60 (\( n=5, 12.5\% \) in both cases). The rating distribution for both are reported (depicted) in fig 1 and fig 2.
Impact of Familiarity and Influence of Current Research on General Acceptance, Willingness to Testify and Reliability for Courtroom Presentation.

Binary logistic regression analysis was applied to analyze the relationship between the predictor variables of expert familiarity and influence ratings of research articles to three different outcome variables: (1) their opinions of general acceptance of eyewitness phenomena, (2) willingness to testify that the phenomena is reliable in a courtroom and (3) whether the phenomena is reliable enough to be presented in a courtroom. For the outcome variable “Please indicate the reliability of the following statement”, for each phenomena, the responses “evidence suggests the reverse is possibly true”, “evidence does not support it”, “evidence is inconclusive”, “evidence tends to favor it” and “I don’t know” were coded as 0 and “evidence is generally reliable” and “evidence is very reliable” were coded as 1. For expert’s willingness to testify and whether the phenomena is reliable enough to be presented in a courtroom, “yes” responses were coded as 1 and “no” responses were coded as 0. The predictors were entered into the model through forced entry.

Results of the binary logistic regression indicated that there was a significant association between influence and familiarity ratings with each of the outcome variables; $\chi^2 (2, 386)= 12.495$, p
<.002 with general acceptance, $\chi^2(2, 386) = 8.606$, $p < .014$ with reliability and $\chi^2(2, 386) = 17.462$, $p < .001$ with willingness to testify. The overall accuracy of classification for general acceptance was 58% ($R^2=.043$), for reliability rating classification was 70% accurate ($R^2=.031$) and 65% for willingness to testify ($R^2=.061$). The model percentage accuracy in classification for general acceptance, reliability and willingness to testify suggests that impact of familiarity and influence on each of the outcome variables is accurate no more than 58%, 71% and 65% respectively.

In each case, familiarity significantly contributed to the model; $p < .010$, $\beta = .116$ and 95% CI [1.028, 1.227] for general acceptance, $p < .049$, $\beta = .098$ and 95% CI [1, 1.216] for reliable enough to be presented in a courtroom and $p < .007$, $\beta = .131$ and 95% CI [1.037, 1.252] for willingness to testify. The results suggest that as familiarity increases by one unit, the odds of phenomena being deemed generally accepted increases by 12.3% (OR = 1.123), reliability increases by 10.3% (OR=1.103) and willingness to testify by 14% (OR=1.140). Influence ratings on the other hand, had a non-significant impact on general acceptance, reliability and willingness to testify ($p < .907$, $p < .855$ and $p < .715$ respectively). Table 5 reports the familiarity and influence ratings for each article. Overall, experts seemed to be more familiar with primary articles than with meta analysis but rated meta-analysis articles more influential on their perceptions of reliability. Additionally, familiarity and influence ratings were highly correlated suggesting that familiarity with an articles had an effect on expert’s appraisal of the article’s influence; $r (396)= .400$, $p < .001$.

**What is Reliable Enough?**

To determine how reliable a topic must be before an eyewitness is willing to testify, each respondent's opinion of a phenomenon was compared with their judgment of whether the phenomenon was reliable enough to be presented in court, and whether they would be willing to testify to the reliability of the particular phenomenon. Table 5 reports the expert opinions for reliability, willingness to testify and scientific bases for opinion for each phenomenon (with Daftary
Overall, experts were consistent in their evaluations. That is, experts were willing to testify 16% of the time when they rated the phenomena as inconclusive (10% in Wave I and 12% in Wave II of Daftary & Penrod, 2012), 87% when generally reliable and 97% when very reliable (97% in Wave I and 98% in Wave II of Daftary & Penrod, 2012).
Discussion

This study set out to (1) assess the extent of general acceptance of various eyewitness phenomena within the scientific community (2) collect information on expert demographics and opinions regarding various other courtroom phenomenon and (3) assess a number of new eyewitness research publications that have come to the fore since the last wave of Daftary and Penrod (2012) survey, which could potentially impact expert’s perceptions of various eyewitness phenomena. Additionally, the survey queried experts about their opinions on various matters involving demographics, research productivity, courtroom involvement and perceptions of the jury’s understanding and consideration of various eyewitness phenomena.

The respondents in the current study were notably less accomplished than those involved in Daftary and Penrod (2012) but seemingly more accomplished than Kassin, et al. (2001) when scholarly achievement through publications were compared. However, courtroom experience was comparable to Daftary and Penrod (2012) with a similar percentage of experts being asked to testify at least once. However, larger percentage of respondents reported that they agreed to testify but fewer percentage of respondents reported that they actually testified in court. A dissimilar trend was noticed when comparing the percentage of times experts agreed to testify in court and the percentage of times they actually testified between Kassin, et al. (1989) findings, Kassin, et al. (2001), and Daftary and Penrod (2012). In previous comparisons, there was a decrease in rates of experts agreeing to testify and increase in rates of them actually testifying. It is possible that experts are either more open to testifying or that their perceptions of phenomena reliability, in light of new meta-analysis (see Daftary & Penrod, 2012) for repeat phenomena, has improved and hence rendered them more open to testifying in court regarding them. There are also a few possible explanations for the decrease in actual testimonies. One explanation could be that experts who participated in the current survey were newer to the field with lesser experience actually testifying. Since courts seem to factor an individual’s previous testifying experience while attempting to
establish expertise (Henderson & Lenz, 2009), some respondents probably did not make the cut and hence did not actually testify. It it also worth noting that many of the respondents reported scheduling conflicts as a reason for being unable to testify. Hence, they probably agreed to testify but may have been unable to due to scheduling conflicts. A more troubling, but equally plausible explanation, could be that though experts are accepting requests to testify at a higher rate, courts are admitting fewer eyewitness expert testimonies (e.g., In Commonwealth v. Selenksi, 2016, the court held that “although it is no longer banned, per se, expert testimony concerning the reliability of eyewitness identification was properly excluded when that identification was not the sole or primary evidence of the defendant’s guilt.” p. 1). Additionally, attorneys might have become more aware of issues surrounding eyewitness accounts and hence seek consultations in more cases. These consultations were probably sought out during discovery phases and the consultations may not ultimately result in court appearances. This could probably be due to the testimony appearing less helpful than expected or due to the emergence of other evidence that renders the testimony less relevant or useful. Another possible explanation for these findings could be that a growing percentage of cases (95% to 97% of criminal cases as reported by Alkon, 2017) involving consultations ultimately end in plea bargains rather than ending up at trial and the consultations with experts may or may not play a role in those plea bargains.

The ratio of criminal and civil requests was also consistent with those reported by Kassin et al. (2001) as was the disproportionate percentage of requests from the defense and prosecution reported by the two previous waves (Darfaty, 2012; roughly twice as many requests for criminal in all three studies). Similarly, the number of requests from the plaintiff in civil cases was considerably higher than requests from the defendant in the current study. Interestingly, in one, the burden of proof lies with the non-requesting party (prosecution) and in the other, it lies with the party pursuing expert testimony (plaintiff).
A majority of experts (93%) reported that the primary role of the expert was to educate the jury. Of those who reported that the expert’s primary role was to provide assistance to a particular party (7%), none of them had actually testified in court. One respondent reported “I wish to avoid 'hired gun' scenarios due to the nature of my research” as a reason for declining to testify suggesting a moral conflict though reportedly believed that the purpose of the testimony was to assist a particular party and another reported that it was their duty to assist the judge as their deciding country did not have a jury system probably referring to the judge as the party they believed the expert’s role was to assist. Moreover, these results were strikingly similar to those collected in 2001 by Kassin, et al. Furthermore, reasons provided for other roles of experts in a courtroom were also consistent (eg: “to educate…judge”). This suggests that the beliefs held by experts on the topic has remained consistent for over 17 years.

Experts were more willing to testify when they deemed a phenomenon as reliable enough and were not willing to testify if they thought the phenomenon was not reliable enough to be presented in court by psychologists. Such testifying behavior was also recorded in the previous surveys. Experts were also noted to be more willing to testify when they deemed a phenomenon generally reliable or very reliable and less likely to do the same when they deemed a phenomenon inconclusive. These findings have remained consistent through the Kassin, et al. survey in 2001, Daftary in 2012 and the in the current study. Furthermore, expert opinions regarding general acceptance, reliability and willingness to testify were highly correlated to the basis of their opinions on published, peer reviewed, scientific research. These findings are important as the admissibility standards set by the courts employing Daubert (Daubert v. Merrell Dow Pharmaceuticals, 1995) require the phenomena in question to have been subjected to peer review and publication and also to have attracted widespread acceptance within a relevant scientific community. Hence, given that eyewitness experts are willing to testify most often when they consider the evidence highly reliable and seem to base their appraisal of reliability on peer reviewed and published scientific information,
this could ease the burden on the judiciary to some extent. Furthermore, this could also help bring some resolution to the question of how reliable a body of research must be before it is presented in the court as experts consistently (in all four surveys) indicate willingness to testify largely when they deem the phenomena in question to be reliable enough for such purposes.

Similar to the experts surveyed in Kassin et al (2001) and Daftaty (2012), experts surveyed in this wave were not influenced by their courtroom experience and research productivity regarding their opinions of phenomena reliability. Also consistent with Kassin (2001), courtroom experience was not an influential factor on expert’s willingness to testify while expert’s research productivity was correlated with their willingness to testify. Specifically, experts with higher research productivity were more willing to testify than those with lower research productivity. A likely explanation for this could be that experts involved with more research were probably more familiar with eyewitness related phenomena and hence were more confident about testifying in court regarding the various phenomena. Experts less involved in eyewitness research were probably less familiar with the phenomena queried and hence less confident and likely to testify regarding the various phenomena. On the other hand, research productivity was not influential on phenomenon reliability ratings (consistent with Kassin, et al., 2001 findings). This suggests that, irrespective of their own courtroom experience and research productivity, there is a general consensus among experts regarding their opinions about eyewitness phenomena and whether it is reliable enough for psychologists to testify to in court. This is useful as courts require this level of general consensus among scientists.

Experts were also queried on topics regarding their opinions about the ken of the jury. Overall, experts did not believe that jury verdicts were influenced by a serious consideration of many of the phenomena queried. Phenomena such as Elderly Witness and High Identification Confidence received above average ratings while the rest of the phenomena received below average ratings. This suggests that experts do not opine that the various eyewitness phenomena examined in
this study are entirely within the *ken* of the jury. On many occasions, such as in *State v. Chapple* (1983), expert eyewitness testimony was excluded as the court ruled that the theories in question were within the *ken* of the jury. Though expert opinions do not establish with absolute certainty the boundaries regarding the *ken* of the jury, it could nevertheless prove helpful as a reference to the courts.

Experts displayed greater consensus regarding acceptance on topics such as Showups, Blind Administration and Change Blindness when compared to the second wave of Daftary and Penrod (2012). Other phenomena such as Lineup Fairness, Elderly Witness, Weapons Focus and Description Matched Lineup did not see much change in consensus. Additionally, none of the experts reported “I do not know” for Presentation Format, Shows, Weapons Focus, Blind Administration, Change Blindness and High Identification Confidence. Furthermore, there was a strong consensus regarding the reliability of various eyewitness phenomena: Showups, Blind Administration, Description Matched Lineup, Weapons Focus, Elderly Witness, High Identification Confidence, Lineup Fairness and Change Blindness. Other phenomena such as Presentation Format, Own Age Bias and Own Gender Bias yielded lower consensus. The current wave also attempted to record eyewitness expert’s appraisal of appropriate effect size that reflects reliable findings in a study for Pearson’s r and for Cohen’s d. Expert’s seemed to reach some consensus on the appropriate Pearson’s r which fell within a narrower range but was more varied on their opinions regarding the appropriate Cohen’s d to reflect reliable findings in a study.

Experts were asked to rate a number of articles regarding their familiarity with the articles and their perceived extent of its influence on their judgements of reliability of the eyewitness phenomena evaluated by the articles. Inconsistent with the findings from the first two waves of Daftary and Penrod (2012), experts were found to be more familiar with primary articles than with meta-analysis. This could probably be attributed to the nature of phenomena addressed by the articles. Most of the primary articles addressed well-researched phenomena while most of the meta-
analyses addressed new phenomena. It is plausible that the repeat phenomena have been a topic of interest among the research community for longer periods of time and hence were more familiar with these articles. However, influence ratings of articles were consistent with the findings from the previous two waves conducted by Daftary and Penrod (2012). It should be noted that overall, meta-analysis articles had higher citation counts when compared to primary articles which may reflect their higher influence ratings.

Unlike the findings form Daftary and Penrod (2012), influence ratings were not influential on expert’s general acceptance of phenomena, reliability ratings or willingness to testify. A possible explanation for this unexpected lack of effect could be that the meta-analysis used in the study, that reflect higher influence ratings, were sufficiently new with regards to publication dates or in topic novelty making it too early for a strong general consensus to have formed among the respondents. Additionally, as the familiarity and influence ratings were highly correlated, it is also possible that this lack of effect was partly driven by the varying levels of familiarity with the topic which could have impacted the influence ratings. Furthermore, this was also reflected in the influence ratings of the individual articles with only two articles, both of which had comparably higher citation counts, scoring an average influence rating above 4.5 on an influence scale ranging from 1 (“not at all”) to 9 (“extensively”). Familiarity on the other hand, seemed to have an impact on expert’s appraisal of general acceptance, reliability ratings of phenomena and willingness to testify. These were also reflected in the familiarity ratings of the individual articles with 4 articles receiving scores above the average 50 on a familiarity scale ranging from 0 (“did not impact my perception of reliability”) to 100 (“significantly impacted my perception of reliability”). These findings were consistent with those from the two waves of Daftary and Penrod (2012).

There are a few limitations of the study, the most notable being a low response rate of 31.95% comparable to the response rate of 34% recorded by Kassin, et al. (2001). In acknowledgement of the low response rate, the list of experts solicited were also quarried about
possible reasons for not participating in the survey. They were asked 3 questions: “I did not participate in the survey because (check all the apply)” with a list of possible reasons: “lack of time”, “length of survey”, “inconvenient timing of participation request”, “do not consider myself an expert int he field”, “plan on participating at a more convenient time”, “not interested” and room for other reasons (2) “I started the survey and chose to discontinue because (enter reason below)” and (3) “what improvements could be made to encourage your participation?”. Of the 10 responses received for (1), 6 respondents did not consider themselves experts in the field, 2 reported the reason as the length of the survey (one of whom further included “too long” as a reason for starting and choosing to discontinue the survey) and 2 reported other reasons that included “I am highly interested. But all the questions are solely relevant in a legal system as the USA…”. Notable reasons for starting and discontinuing included “I lack confidence...needed to brush up and check on the latest findings…” and “Many of the questions referred to recent findings and I was not qualified to give valid responses.”. These responses suggest that experts acknowledged the need for expertise on the topics surveyed, wanted to provide fully informed responses and those that were unable to do so probably refrained from participation. Additionally, they were asked for suggestions to encourage participation and respondents suggested similar reasons as those for not participating such as the length of the survey and clarification regarding participant confidence levels “ How confident do you want your respondents to be?”. Altering the length of the survey is a challenging as all questions included on the survey are relevant and important to for the aims of the study. Confidence on the other hand could be problematic to define in the context of expertise. High confidence levels might discourage participation and low confidence levels might yield inaccurate responses. Furthermore, perceived self-confidence might not be an actual representation of knowledge and expertise.

The experts surveyed were not a comprehensive list of all experts in the field. However, an attempt was made to identify researchers who are currently active for solicitation of responses to the
survey in addition to those who participated in the previous two waves of Daftary and Penrod (2012). It is also plausible that some experts in the field were not contacted via the methodology used in this study. Additionally, all the data collected by the survey are based on self-report leaving room for the possibility that some respondents may have over-estimated or even under-estimated certain occurrences such as their research publications, courtroom experience, familiarity with and influence of the literature surveyed. The scope of this study was restricted to expert testimony regarding eyewitness phenomena. Research into similar expert consensus in other domains of psychological phenomena (such as expert testimony on earwitness, battered woman syndrome, rape trauma, etc.) would also make similar contributions. Such research could also assess the sources of influence on expert opinion in the respective fields. As experts were disproportionately involved in criminal and civil cases, future research could also look more closely into the role of eyewitness expert testimony in both cases.

Conclusion

The present study revealed some important consistencies and some changes in the opinions of eyewitness experts. Expert’s perception and management of their activity in the courtroom indicated that they set relatively high standards for their own involvements as respondents were willing to testify mostly when they deemed a phenomena as "generally reliable" and "very reliable.” Furthermore, reliability and willingness to testify seemed to be independent of respondent expertise level while expert’s willingness to testify was influenced by their research productivity. High levels of agreement amongst experts on a variety of topics such as Showups, Blind Administration, Description Matched Lineup, Weapons Focus, Elderly Witness, High Identification Confidence, Lineup Fairness and Change Blindness was also recorded. Familiarity with various eyewitness research publications seemed to have an effect on expert’s perception of general acceptance, assessments of reliability of eyewitness phenomena and willingness to testify in court. The present results should provide needed guidance to judges in their decision as gatekeepers, the courts in
seeking information on the acceptance of novel research areas, lawyers in their examination of
eyewitnesses and psychological experts in determining which phenomena are reliable enough to
present in court and also in understanding the sources that influence consensus formation. This
should alleviate some of the issues surrounding the admission of expert testimony in the courtroom.
References


United States v. Holloway, 971 F.2d 675 (11th Cir. 1992)


