2016

Meta-Analysis as an Aid for Judicial Decision Making

Nancy K. Steblay
Augsburg College in Minneapolis

Meta-Analysis as an Aid for Judicial Decision Making

Nancy K. Steblay

Judges are the gatekeepers for courtroom evidence. A decision about scientific-evidence admissibility requires knowledge of the relevant content and a judgment about the quality of what is often an enormous and complex research literature. How are judges to learn about a highly specialized topic? One option is to immerse oneself in the scientific research. Alternatively, judges may read legal briefs for related cases, attend judicial-education seminars, or open their courtrooms to experts at pretrial hearings or at trial. In any of these ventures, one is likely to run into the research review technique called *meta-analysis*.

This article has two primary and parallel objectives: (1) to describe what one can expect to find within a high-quality published meta-analysis, should one choose to go directly to the research literature, and (2) to propose how the qualities of a good meta-analysis can serve as a standard for how scientific knowledge is transferred to judges and juries when scientific experts become a part of the legal process (*e.g.*, via judicial training, pretrial hearings, or expert testimony).

EYEWITNESS SCIENCE AS AN EXAMPLE OF META-ANALYTIC REVIEW

One specific research realm—eyewitness science—can provide an illustration of how meta-analytic reviews can assist the court in its charge to make sense of a vast research literature and to render scientific information useful to triers of fact. A consideration of meta-analysis via eyewitness science is timely. In a recent *Court Review* article, Smarlarz and Wells alerted trial judges to the fact that rapid scientific advances in eyewitness-identification evidence have created a new burden for gatekeepers to the evidence permitted in court.[1]

Recent legal developments regarding the evaluation of eyewitness evidence have prompted changes for jury instructions and pretrial hearings. The New Jersey Supreme Court produced an expansive summary of eyewitness-memory research in a Special Master's Report that guided its decision in *State of New Jersey v. Henderson*.[2] The Oregon Supreme Court likewise took on the massive eyewitness literature to address the fundamental problem of mistaken eyewitness identification. The Oregon Court's science-based analysis in *State of Oregon v. Lawson* ultimately repositioned eyewitness evidence to align with state evidence law.[3] The Court's decision cited the source of probative value in the *original* memory of the eyewitness, uncontaminated by outside information. Hence, the ruling requires Oregon judges to scrutinize impending eyewitness testimony to see that it is based on the witness's personal perception and knowledge—regardless of whether or not law enforcement used a suggestive identification procedure. Implicit in this ruling is that judges can decipher the complex findings of eyewitness science and apply them in a meaningful way.

On the heels of the *Lawson* and *Henderson* decisions, the National Academy of Sciences (NAS) was asked in 2012 to critically assess the status of eyewitness-memory research and to offer advice and recommendations where appropriate. This society of distinguished scholars, established by an Act of Congress and signed into law by President Abraham Lincoln in 1863, is charged with providing independent objective advice to the nation on matters related to science and technology. The National Research Council of the NAS found the body of eyewitness research to be rigorous, sound, and compelling in its importance to the justice system. The Council followed with a set of actionable recommendations for strengthening the integrity of eyewitness evidence collected by police and brought to court. Most relevant to this audience is the National Research Council's position that judges hold an affirmative obligation to ensure the reliability of eyewitness evidence presented at trial. The expectation is that judges will transfer this information to a jury in the form of *scientific-framework expert testimony* and clear and concise jury instructions.[4]

The problems inherent to eyewitness memory are not new to the legal system. Eyewitness reliability has been questioned repeatedly over decades, and attempts to tamp down the potential negative impact of faulty memory have been considered at all levels of courts, including the U.S. Supreme Court—most recently in *Perry v. New Hampshire*.[5] What may be new to many in the legal system, however, are the specific findings of psychological scientists on eyewitness-memory issues. In part to aid legal understanding of what is now a mature social-science literature, eyewitness-memory scientists have produced a series of meta-analytic reviews. Meta-analysis provides a foundation for the scientific framework that experts can bring to the court. A small set of these reviews are used below to illustrate the quantitative review technique of meta-analysis.

Footnotes

1. Laura Smalarz & Gary L. Wells, *Eyewitness-Identification Evidence: Scientific Advances and the New Burden on Trial Judges*, 48 Ct. Rev. 14 (2012).
2. 27 A.3d 872 (N.J. 2011); see also Report of the Special Master, *available at* http://www.judiciary.state.nj.us/pressrel/HENDERSON%20FINAL%20BRIEF%20.PDF%20(00621142).PDF.
3. 352 Or. 724 (Or. 2012).
4. National Research Council of the National Academies, Identifying the Culprit: Assessing Eyewitness Identification 74-77 (2015), *available at* http://www.nap.edu/openbook.php?record_id=18891.
5. 132 S. Ct. 716 (2012).

THE RATIONALE FOR META-ANALYSIS

The phrase *meta-analysis* simply means "analysis of analyses." Quantitative results (statistics) from a set of independent studies that have all tested the same proposition are combined into one overall statistical analysis. The purpose is to determine what the overall pattern of the data looks like—across studies, from many labs, from different participant samples, and from a variety of experimental protocols and measures. The meta-analyst does the work of reading the available studies on a topic, combining the findings, and producing a meaningful synthesis of results.

The rationale for meta-analysis begins with five basic tenets of the scientific method.[6]

1. No single study offers a complete or final answer. A single test of a hypothesis cannot include all possible participants, laboratory stimuli, and relevant measures. A single eyewitness study will test, for example, a defined sample (200 undergraduates), a unique portrayal of an event (a 90-second video of a carjacking), and a specific identification procedure (a six-photo lineup of Caucasian men). Each study is inherently confined to a limited version of the phenomenon of interest.
2. So, more is better. Good science is anchored in the rule that a study must be exposed to independent testing and replication by other scientists. As independent studies cumulate, we are able to see whether the research outcome that appeared in one lab can be reproduced by other scientists even as these researchers use somewhat different experimental techniques, materials, and participant samples.
3. Converging evidence is powerful. Scientific evidence becomes more persuasive as study results repeatedly point in the same direction. In applied science, convergence between laboratory and field tests is particularly compelling.
4. Study results will vary. Not surprisingly, the use of many different samples of participants, materials, and procedures will produce some variation in outcomes. This variability is expected and informative. Scientists often tinker with a phenomenon to determine the conditions under which an established scientific principle will remain stable or begin to break down.
5. Knowledge is cumulative. Every study (assuming decent-quality methodology) offers some information, a piece of a developing mosaic for our knowledge about a phenomenon.

CORE COMPONENTS OF A META-ANALYSIS

A meta-analysis offers a synthesis of scientific research on a specific topic, with three core components[7]—and one additional highly desirable attribute that I describe below.

A Statistical-Significance Test. This refers to the overall support for the hypothesis (the experimental treatment) across studies: Is this hypothesis supported to a level of scientific certainty? (Can we trust this outcome?) This is stated as a statistic with an attached probability-value ("*p-value*"), a number between 0 and 1.00 that describes the likelihood that the observed results would have occurred by chance (a fluke) if there were no true difference between the experimental conditions. A *p-value* smaller than .05 is the traditional level for "significant results" and noted in words such as "this difference between groups is statistically significant."

A meta-analysis offers a synthesis of scientific research on a specific topic . . . .

For example, a meta-analysis of 16 published experiments involving 3,196 adult witness-participants reported that using a cautionary lineup instruction to an eyewitness ("the person you saw commit this crime may or may not be in this lineup") can significantly reduce eyewitness misidentification errors. The likelihood is small that this comparative difference in eyewitness errors (between witnesses who heard the instruction and witnesses who did not) would have occurred in these studies if there were no real difference ($p < .05$). In short, we trust that this phenomenon is real within the bounds of scientific certainty. A cautionary instruction makes a difference for eyewitness identification accuracy.[8]

An (Average) Effect Size. An effect-size statistic addresses a different and important question: What is the *size* of the difference between tested groups? Or, just how *strong* is the relationship between the measured variables? There are a number of favored statistical indicators of effect size (among them, *r*, *d*, *h,* and the *odds-ratio*, *OR*) and rules of thumb for interpreting each. But the basic notion is that the effect size indicates "how much?" or "how strong?" An excellent guide to the calculation of meta-analytic statistics for legal scholars is provided by Blumenthal.[9]

Hence, for example, a cautionary lineup instruction significantly reduces eyewitness identification errors, a good thing. But how large is the reduction? The meta-analysis of laboratory studies reported an average 27% fewer errors when witness-participants were given the cautionary instruction. Conversely stated, the witnesses who were *not* cautioned produced 27% more identification errors. This percentage difference is a direct measure of effect size, but it can also be reported as $h = .63$ or $r = .31$. In any case, the rule-of-thumb translation is that this is a medium-size effect.[10]

Moderators of the Effect. Studies that test the same principle will produce some variation in effect sizes. Often a meta-analysis is driven by the fact that there seems to be inconsistent outcomes among studies. That is, sometimes the hypothesis is supported, sometimes it is not supported and may even be contradicted. Moderator analyses can help resolve theoretical, method-

6. *See generally* KEITH E. STANOVICH, HOW TO THINK STRAIGHT ABOUT PSYCHOLOGY (2004).
7. *See generally* Robert Rosenthal, *Meta-Analytic Procedures for Social Research*, *in* 6 APPLIED SOCIAL RESEARCH METHODS SERIES (Leonard Bickman & Debra J. Rog eds., 1991) (discussing the evolution of meta-analysis procedures).
8. Nancy K. Steblay, *Lineup Instructions*, *in* REFORM OF EYEWITNESS IDENTIFICATION PROCEDURES 65 (Brian L. Cutler ed., 2013).
9. Jeremy A. Blumenthal, *Meta-Analysis: A Primer for Legal Scholars*, 80 TEMP. L. REV. 201, 201 (2007).
10. *Id.*

A good meta-analysis will . . . place the research within a context that includes the *quality* of the research methodology . . . and the *relevance* of the information . . . .

ological, or practice questions by figuring out why variations in study outcomes occur.

Moderators are variables (factors) that influence the size of an effect. Often moderators are aspects of how the study was conducted—decisions made by a researcher about who the participants were and what they experienced. For example, a recent meta-analysis documented the moderating effect of eyewitness age on lineup identification accuracy: Older eyewitnesses (60+ years of age) are more than twice as likely to make an identification error than are younger eyewitnesses (18-25 years), an *odds-ratio* (*OR*) that exceeds 2.0.[11] Also of interest, eyewitness age moderates the benefit of a cautionary lineup instruction; the instruction is often lost on older adult eyewitnesses.[12]

The forensic relevance of a research finding can be established in part by analysis of moderators: whether the effect is persistent despite changes in sample or method, or conversely, if the effect is limited or even absent under conditions that are meaningful to real-world applications. An important benefit of meta-analysis is to examine just how far an effect extends.

So What? The Context for the Research Outcomes. The statistics of a meta-analysis will provide answers to core questions about the scientific status of a research finding: Is the outcome reliable by scientific standards? Is the effect small, medium, large? Are there significant moderators of the effect? But the follow-up question is crucial: Is this research finding *important*? Particularly for applied science, the interpretation of a significance test and an effect size—the "story" told by the statistics—rests with the informed judgment of the researcher as well as the policymakers or the triers of fact who must apply this research to a question at hand.

A good meta-analysis will not only report statistical results and moderator analyses but also place the research within a context that includes the *quality* of the research methodology (including strengths or gaps in the extent knowledge and the developmental stage of the research) and the *relevance* of the information to theory and practice. In short, the challenge for meta-analysis is to find and report the *knowledge* in all that information.[13]

Research Quality

All tests, published and unpublished, are typically included in meta-analytic calculations so that the analyst can work with an increased amount of information and establish whether the published studies are an anomalous subset of the broader research. At the same time, however, the *Daubert* criteria (from *Daubert v. Merrill Dow Pharmaceuticals*)[14] and other admissibility standards create a compelling reason to examine only published peer-reviewed work as a means to ensure decisions are based on high-quality vetted research.

There are various approaches to the assessment of quality in meta-analysis. Many meta-analytic projects involve a substantial effort to locate unpublished work ("file drawer" studies) by contacting researchers and combing through convention programs and Internet sites for unpublished work. The status of an article as published or unpublished can then be used as a moderator to test whether unpublished studies produce a different outcome and, if so, why that might be. However, when unpublished and published work show different outcomes, it is difficult to interpret this finding. This is because unsound methodology may have kept the unpublished work out of peer-reviewed journals. With some exceptions (newer unpublished studies may ultimately move into scientific journals), unpublished studies may involve problems of small sample size, unrefined pilot projects, lack of experimental controls, or methodological details long forgotten and thus unavailable for peer review. These shortcomings may contribute to the experimental effects, or lack thereof, obtained in unpublished work.

Many meta-analyses do not include unpublished work and instead calculate a *fail-safe-n* that estimates the number of non-supportive unpublished tests that must exist in the file drawers to overturn a statistically significant finding in the published work.[15] The report of a *fail-safe-n* is quite straightforward (*e.g.*, "20 non-supportive studies could negate this finding"), and the informed judgment of the researcher will help to estimate the likelihood that such a quantity of non-supportive unpublished studies could reasonably exist in the file drawers.

Methodological quality must be addressed even within published studies. To do so, each individual study can be weighted in the summary calculations by a factor of quality. The most common weight is sample size, a strategy that rests on the reasonable assumption that larger samples will better represent the population from which the sample was drawn and lead to more stable and accurate results and thus should contribute more heavily to the summary statistics. Of course, this technique would weight a weakly designed study of 1,000 participants more heavily than a well-designed study of 100. Hence, there are further quality considerations: the adequacy of experimental design and measures used, the authenticity of the experimental manipulations, the completeness of the reported data, the appropriateness and rigor of the statistical tests. Each of these methodological factors may become a criterion for including or excluding a study in the first place or may become the basis for a moderator test. Whatever the case, a competently conducted meta-analysis will clearly state up front the criteria for including/excluding studies and clear operational

11. Ryan J. Fitzgerald & Heather L. Price, *Eyewitness Identification Across the Lifespan: A Meta-Analysis of Age Differences*, 141 PSYCHOL. BULL. 1228 (2015).
12. Rachel A. Rose, Raymond R. Bull & Aldert Vrij, *Non-Biased Lineup Instructions Do Matter: A Problem for Older Witnesses*, 11 PSYCHOL. CRIME & L. 147 (2005).
13. Gene Glass, *Primary, Secondary, and Meta-Analysis of Research*, 5 EDUC. RESEARCHER 3 (1976).
14. 509 U.S. 579 (1993).
15. *See* Rosenthal, *supra* note 7.

definitions for each factor considered in the analysis and will then follow these exacting standards.

Some realms of research may require a more nuanced approach to unpublished data. There is a meaningful difference between methodologically flawed unpublished research and sound but unpublished data that could bring forth null or contradictory results. The latter are informative and require appropriate viewing and evaluation by a research community. For example, in tests of medical devices or pharmaceuticals or treatment outcomes, unpublished but high-quality test trials can contribute essential information to fully evaluate scientific claims.

Relevance

The usefulness of a meta-analysis is, of course, limited by the number of available individual studies on which it is based and by the depth and breadth of those research products. In the early stages of research, a simple meta-analysis can speak to the basic causal or correlational relationships between variables as revealed in a small extant literature, thereby providing a foundation and direction for future studies. A later meta-analysis on the same topic may include lab and field tests, a greater diversity of participants, and theoretical or methodological refinements that extend the research in new directions. It is not unusual to see multiple meta-analyses on the same topic that trace an increasing volume and complexity in the research findings over time.

For applied legal research, a meta-analysis may be able to empirically address questions that directly speak to police procedure, policy, or legal evidence. For example, the cautionary instruction to eyewitnesses is among a number of recommended lineup procedural reforms that try to tamp down the dangerous tendency of many eyewitnesses to make an identification even when the guilty perpetrator is not in the lineup (a "culprit-absent" lineup). The risk, of course, is to an innocent suspect when the true culprit is not even in the lineup, as has been the case with hundreds of DNA exonerees highlighted from work of The Innocence Project.[16] The unbiased lineup instruction is useful to reduce witness choosing from a culprit-absent lineup by an average 27%. But what happens when witnesses view a lineup that *does* include the guilty culprit? The cautionary instruction reduces correct culprit identifications by 5% (a small effect) as well as reducing mistaken identifications.[17]

A policy recommendation for a cautionary instruction with police lineups will weigh the costs and benefits of the procedure, including the fact that police lineups are very common, thousands are conducted each year with an unknown subset involving innocent suspects. Lineups are much more proficient at incriminating the guilty than they are at exonerating the innocent. Even a small percentage increase in errors translates to a large number of misdirected investigations and dangerous mistaken identifications. Also, failure to use a cautionary instruction may be exacerbated by other police practices that place innocent suspects at risk. A thoughtful consideration of this complex integration of research and policy considerations has recently been provided by Wells, Yang, and Smalarz.[18]

[A] meta-analysis may be able to empirically address questions that directly speak to police procedure, policy, or legal evidence.

A judge or jury will retroactively consider whether the lack of a cautionary instruction may have increased the likelihood that the defendant was erroneously chosen from a police lineup—especially in combination with a biased lineup construction, a "non-blind" lineup administrator, and other biasing lineup factors. Conversely, of course, a jury may consider the ramifications of a very sound lineup procedure with a fair lineup structure, blind administration, and proper instruction.

AN ADDITIONAL EXAMPLE: THE POST-IDENTIFICATION FEEDBACK EFFECT

Eyewitness identification procedures have drawn substantial attention over the past three decades as the number of DNA-exoneration cases has grown. More than 70% of these wrongful convictions have been found to involve eyewitness identification error.[19] Garrett's 2012 analysis of DNA-exoneration cases indicated that up to 57% of mistaken witnesses who testified confidently at trial had been quite uncertain at the initial identification: 40% did not identify the defendant on the first try, 21% admitted uncertainty, and 9% said they did not see the face. Even so, witness confidence grew over time, culminating in convincing trial testimony that helped to convict an innocent defendant.[20]

A core concern for eyewitness evidence is that witness confidence about a lineup identification is likely to inflate as the witness learns that he or she chose the police suspect, that another witness also chose that lineup member, that the suspect has been arrested and charged, or that the case against the defendant is strong. The cumulative result is a confident witness on the stand, even if the identification was mistaken. Eyewitness researchers have studied this tendency for witness confidence to grow by examining the slice of time immediately after the lineup identification is made. Is it possible that a well-intended simple comment of confirming feedback from the lineup administrator ("Good job! You identified the suspect.") can trigger witness confidence inflation?[21]

The typical laboratory test for post-identification feedback

16. The Innocence Project, http://www.innocenceproject.org.
17. Steblay, *supra* note 8.
18. Gary L. Wells, Yueran Yang & Laura Smalarz, *Eyewitness Identification: Bayesian Information Gain, Base-Rate Effect Equivalency Curves and Reasonable Suspicion*, 39 Law & Hum. Behav. 99 (2015).
19. *Eyewitness Misidentification*, The Innocence Project, http://www.innocenceproject.org/causes/eyewitness-misidentification/.
20. *See* Brandon L. Garrett, Convicting the Innocent: Where Criminal Prosecutions Go Wrong 233, 285-87 (2011).
21. Nancy K. Steblay, Gary L. Wells & Amy Bradfield Douglass, *The Eyewitness Post Identification Feedback Effect 15 Years Later: Theoretical and Policy Implications*, 20 Psychol. Pub. Pol'y & L. 1 (2014).

[T]his strong literature provides converging evidence for other recommendations from eyewitness scientists . . . .

(more simply called a *feedback effect*) is as follows: Witness-participants view a crime event (*e.g.*, video clip of a robbery) and then attempt to identify the culprit from a lineup. Witnesses are randomly assigned to receive confirming feedback ("Good, you identified the suspect.") or no feedback after their identification. Later, the witness is asked to retrospectively report a number of testimony-relevant judgments, including certainty in the identification, attention paid and quality of view at the time of the crime, and additional aspects of eyewitness experience (trust in witness memory, willingness to testify). The clincher in this experimental design is that all witnesses have unknowingly made an identification error from a lineup that did not include the culprit; all have selected a known-innocent lineup member.[22]

Judges and attorneys who recognize the relevance of this scenario for potential problems in eyewitness testimony—and the danger to an innocent defendant—will find dozens of articles on the topic. How to synthesize these resources? A meta-analysis has done the work.[23] Steblay, Wells, and Douglass found 23 published articles from 11 separate laboratories, representing 7,000 witness-participants from the United States, Canada, Europe, and Australia.

The post-identification feedback effect is one of the strongest and most robust effects in the eyewitness literature. The meta-analytic results include the following:

A Statistical-Significance Test for the Overall Pattern. The primary measure of interest is the witness's retrospective certainty about the identification decision ("How certain were you at the time of the identification?") Witnesses who received confirming feedback expressed significantly more certainty about their identification decision compared to witnesses who received no feedback ($p < .0001$). The impact of the lineup administrator's confirming feedback on eyewitness confidence in a mistaken identification is established with scientific certainty.

The Size of the Effect. How big was the impact of this simple comment from the lineup administrator ("Good, you identified the suspect.") on witness confidence? *Effect size* is presented as a statistic ($d = .98$), a measure of how far apart the certainty scores of the two compared groups are (the feedback vs. no-feedback groups). Readers conversant with statistics will understand these results: The feedback inflates the mean (average) certainty of eyewitnesses by a full standard deviation. In simple rule-of-thumb terms, the effect size is very large.

Moderators of the Effect. The impact of confirming feedback stretches across an array of eyewitness judgments, including certainty and ease of making the identification, belief in a strong memory basis for the ID, willingness to testify in court, and retrospective judgments about how good the view was and how much attention was paid to the crime event. However, the feedback effect does not extend to witness judgments of how long (seconds, minutes) they viewed the crime and at what distance. Feedback appears to distort witness retrospective judgments about subjective experience ("His photo just popped out at me!") but not objective physical measures.

The passage of time between identification and feedback does not weaken the influence of feedback, and the effect occurs using even generic feedback ("You've been a really great witness."). Children, older adults, and ear-witnesses are also vulnerable to feedback, as are actual witnesses to serious crimes in ongoing investigations.

So What? The Context for the Research Finding. Statistical outcomes are generally rather dry and uninformative for lay audiences. A challenge is to communicate this information in a forensically relevant way. One way to think about the impact of confirming feedback is this: the percentage of *mistaken* eyewitnesses who will display high certainty (at a confidence level of 80% or higher) rises from a mere 6% (when there is no feedback) to 29% (with feedback)—a nearly fivefold increase. This is a dramatic escalation in *highly confident but mistaken* eyewitnesses, who also are now convinced that they paid very good attention and had a good view of the crime. Post-identification feedback distorts eyewitness testimony and contaminates eyewitness evidence.

This meta-analysis is the second conducted on the feedback effect; the first was a decade earlier in 2006.[24] The increased number and scope of studies allowed the newer meta-analysis to report the discouraging results of attempted interventions to dispel the feedback effect. Once the feedback is provided, it is difficult if not impossible to "un-ring that bell." The authors now offer the recommendation that feedback effects must be avoided rather than remediated and that the best means is through use of a "blind" lineup administrator who has no knowledge of who the suspect is in the lineup. Furthermore, there is also good reason to document the confidence and testimony-relevant judgments at the time of the identification before any form of feedback.

For policy consideration, this strong literature provides converging evidence for other recommendations from eyewitness scientists regarding double-blind lineup administration, securing a confidence statement at the time of the ID, and videotaped lineup procedures. The objective of these recommendations is to obtain a clear record of witness statements before the introduction of the inevitable memory contaminants that come from the witness's subsequent experiences.

The post-identification feedback effect is also particularly relevant for consideration of judges and triers of fact. There is a need for prosecutors, judges, and juries to establish whether disparity exists between a witness's report at the time of the identification and a later report after witness exposure to external influences. To this point, the research indicates that eyewitnesses typically do not recognize the impact of the feedback on their judgment, so this confidence distortion carries over to their testimony. The feedback can even impair witnesses' mem-

22. *Id.*
23. *Id.*
24. Amy Douglass & Nancy K. Steblay, *Memory Distortion in Eyewitnesses: A Meta-Analysis of the Post-Identification Feedback Effect*, 20 Applied Cognitive Psychol. 859 (2006).

SOME META-ANALYSES OF INTEREST TO JUDGES

THE DETECTION OF DECEPTION:

Charles F. Bond, Jr. & Bella A. DePaulo, *Accuracy of Deception Judgments*, 10 PERSONALITY & SOC. PSYCHOL. REV. 314 (2006), *available at* https://www.researchgate.net/publication/6927452_Accuracy_of_Deception_Judgments.

NONVERBAL INDICATORS OF DECEPTION:

Siegfried L. Sporer & Barbara Schwandt, *Moderators of Nonverbal Indicators of Deception: A Meta-Analytic Synthesis*, 13 PSYCHOL. PUB. POL'Y & L. 1 (2007), *available at* http://dx.doi.org/10.1037/1076-8971.13.1.1.

RISK ASSESSMENT:

Min Yang, Stephen C.P. Wong & Jeremy Coid, *The Efficacy of Violence Prediction: A Meta-Analytic Comparison of Nine Risk Assessment Tools*, 136 PSYCHOL. BULL. 740 (2010), *available at* http://www.ncbi.nlm.nih.gov/pubmed/20804235.

POLICE INTERVIEW STRATEGY:

Amina Memon, Christian A. Meissner & Joanne Fraser, *The Cognitive Interview: A Meta-Analytic Review and Study Space Analysis of the Past 25 Years*, 16 PSYCHOL. PUB. POL'Y & L. 340 (2010), *available at* http://works.bepress.com/christian_meissner/57/.

COMPETENCY TO STAND TRIAL:

Gianni Pirelli, William H. Gottdiener & Patricia A. Zapf, *A Meta-Analytic Review of Competency to Stand Trial Research*, 17 PSYCHOL. PUB. POL'Y & L. 1 (2011), *available at* https://www.apa.org/pubs/journals/releases/law-17-1-1.pdf.

JUDICIAL INSTRUCTION TO DISREGARD INADMISSIBLE EVIDENCE:

Nancy Steblay, Harmon Hosch, Scott Culhane & Adam McWethy, *The Impact on Juror Verdicts of Judicial Instruction to Disregard Inadmissible Evidence: A Meta-Analysis*, 30 LAW & HUM. BEHAV. 469 (2006), *available at* https://www.researchgate.net/publication/6881940_The_Impact_on_Juror_Verdicts_of_Judicial_Instruction_to_Disregard_Inadmissible_Evidence_A_Meta-Analysis.

PRETRIAL PUBLICITY:

Nancy Steblay, Jasmina Besirevic, Solomon Fulero & Belia Jimenez-Lorente, *The Effects of Pre-Trial Publicity on Juror Verdicts: A Meta-Analytic Review*, 23 LAW & HUM. BEHAV. 219 (1999), *available at* https://www.researchgate.net/publication/226920798_The_Effects_of_Pretrial_Publicity_on_Juror_Verdicts_A_Meta-Analytic_Review.

CONFESSION EVIDENCE:

Christian A. Meissner et al., *Accusatorial and Information-Gathering Interrogation Methods and Their Effects on True and False Confessions: A Meta-Analytic Review*, 10 J. EXPERIMENTAL CRIMINOLOGY 459 (2014), *available at* http://works.bepress.com/christian_meissner/60/.

JUROR AND DEFENDANT CHARACTERISTICS:

Dennis J. Devine & David E. Caughlin, *Do They Matter? A Meta-Analytic Investigation of Individual Characteristics and Guilt Judgments*, 20 PSYCHOL. PUB. POL'Y & L. 109 (2014), https://www.researchgate.net/publication/262911046_Do_They_Matter_A_Meta-Analytic_Investigation_of_Individual_Characteristics_and_Guilt_Judgments.

ories of the perpetrator when they see that person later.[25] Furthermore, when participant-witnesses in the lab are asked to observe and judge the accuracy of eyewitness testimony, mistaken witnesses who have received confirmatory feedback are likely to be believed at rates equal to accurate eyewitnesses. In short, confirmatory feedback eliminates the evaluators' ability to discriminate between accurate and mistaken eyewitness testimony.[26] Hence, as the Oregon Supreme Court noted in *State of Oregon v. Lawson*, the danger of confirming feedback "lies in its potential to increase the *appearance* of reliability without increasing reliability itself."[27] The *Lawson* decision highlighted the elasticity of witness certainty and the problems for eyewitness evidence when witness confidence in memory is overstated. The implication is that subsequent trial testimony of the witness will portray a misleading level of certainty and distorted reports of the witness's actual experience.

The feedback-effect meta-analysis provides a powerful statement that is relevant to law and judicial decision making. The research strikes at the heart of the U.S. Supreme Court's 1977 *Manson v. Braithwaite* test for eyewitness reliability used in many suppression hearings.[28] The commonsense assumptions of the court—that were articulated before scientific exploration of eyewitness issues—are now challenged by eyewitness scientists. Three of the five *Manson* criteria (witness certainty at the time of identification, witness opportunity to view the offender, and witness degree of attention to the culprit) are all based on witness self-reports that can be easily distorted by the suggestion from confirmatory feedback.[29]

LINKING META-ANALYSIS AND SCIENTIFIC-FRAMEWORK KNOWLEDGE: IMPLICATIONS FOR JUDGES AND THE COURTROOM

What can one expect from a meta-analysis (or from expert-provided legal training)?

Clarification: The Cut-to-the-Chase. The purpose of a meta-analysis is to clarify a body of research findings—to make it comprehensible—to say, in essence, this claim holds up, or it doesn't, or it holds up only under certain conditions. To be

25. Laura Smalarz & Gary L. Wells, *Confirming Feedback Following a Mistaken Identification Impairs Memory for the Culprit*, 38 LAW & HUM. BEHAV. 283 (2014).
26. *Id.*
27. 352 Or. 724 (Or. 2012).
28. 432 U.S. 98 (1977).
29. *See generally* Gary L. Wells & Deah S. Quinlivan, *Suggestive Eyewitness Identification Procedures and the Supreme Court's Reliability Test in Light of Eyewitness Science: 30 Years Later*, 33 LAW & HUM. BEHAV. 1 (2009) (discussing the shortcomings of the *Manson* criteria).

A meta-analysis offers a type of prophylactic against the temptation to cherry-pick a single study to champion a position.

sure, some meta-analytic reports can be dense and confusing, especially when the analyses examine theoretical nuances or complex interactions among variables (the "why" of a phenomenon). Nonetheless, a meta-analysis will (or should) provide a clear summary statement about the empirical status of a core hypothesis.

Scientific-Framework Perspective. The notion of "scientific-framework evidence" fittingly describes a good meta-analysis, in that the meta-analysis should provide a clear framework for organizing multiple independent studies and seemingly disparate information. It will help the reader to *think* about the knowledge, not simplistically, but in a more comprehensive and effective way.

A meta-analysis offers a type of prophylactic against the temptation to cherry-pick a single study to champion a position. Simply put, a reliable pattern across studies is stronger evidence than any single study. Of particular relevance to trial judges is that one should not be distracted or misled by a single study or by an expert who cites a single study as definitive. A single study—whether in support of a hypothesis or touting contradictory results—should not carry disproportionate weight.

Yet it is not uncommon for an expert to use a single study as a vivid illustration. This can make good sense: Audiences (and juries) can often better understand the story-like narrative that flows from a single study and its outcome than an unending string of statistical summary statements. But, and this point is critical, the expert's single-study-as-example must be representative of the broader reliable research base. Hence, a qualified expert should know and be able to report extant meta-analytic results for a given topic *and* be able to faithfully represent that literature with an example of a study that is informative and understandable to triers of fact.

Probabilistic Evidence: Meta-Analytic (Science) Findings Are Probabilistic, Not Absolute. Many years ago, the term *social framework evidence* for legal decisions was coined by Monahan and Walker[30] to describe the use of social-science research findings as a context for triers of fact as they adjudicate a specific case. Social (or science) framework evidence does not bear directly on the case facts to be decided by a jury but instead provides the psychological or social-context principles to help triers think about and evaluate claims that do bear on their ultimate decision.

Meta-analysis as scientific framework cannot offer the ultimate answer for a specific case (Did *this* witness make a false identification? Did this product harm the plaintiff?). Rather, it offers a statement regarding the general probability of a relevant proposition, perhaps most easily understood as a statement of risk ("Under this condition, there is a significantly increased risk for misidentification." "Use of this drug significantly increases risk of heart attack."). At the same time, reliable scientific findings can alert triers of fact to mistaken ideas and common myths that are better replaced with scientific knowledge. To wit: memory is *not* like a video-recorder; witness confidence on the stand is *not* a reliable indicator of accuracy.

This point extends to established legal doctrine. Smalarz and Wells cite apparent weaknesses in judicial understanding of eyewitness research.[31] The research literature shows "a tendency for conventional legal understandings (a) to fail to appreciate the power of suggestive procedures, (b) to rely too much on eyewitness-identification certainty, (c) to have faulty views of factors that impair memory, and (d) to generally fail to create disincentives for suggestive procedures."[32] The National Academy of Sciences report is even more strongly worded: "The best guidance for legal regulation of eyewitness identification is not from Constitutional rulings, but from the careful use and understanding of scientific evidence to guide fact-finders and decision-makers."[33]

Limits to the Research Findings: What Am I Not Seeing? When studies are combined in a meta-analysis, one begins to see the gaps in the overall program of research. Thus, the meta-analyst (and expert) should be equipped to speak about strengths and weaknesses in the knowledge base. Most notably, the meta-analysis should expose moderators that dilute or extinguish an effect.

An effective scientific-framework presentation will link foundational laboratory studies with relevant field tests to address the reasonable question about how a lab-based scientific principle will play out in the complicated real world. In some research realms, this is a less pressing question; for example, lab studies can effectively establish the limits for visual perception under varying levels of distance and illumination. For other topics, the connection between lab and field tests is more pertinent. Hence, for example, increased stress of a crime event negatively impacts eyewitness lineup-identification accuracy in lab studies. It is not ethical to subject laboratory participants to high levels of stress or violence. Nevertheless, it is reasonable to speculate that the trajectory of these lab results will hold outside the lab. Field tests in fact confirm that eyewitnesses experiencing very high levels of stress in real conditions (*e.g.*, military personnel survival training; police involved in crime simulations) show reduced identification accuracy.[34]

The Challenge of Connecting Multiple Lines of Research. One key limitation of a meta-analysis is that it addresses one

30. John Monahan & Laurens Walker, *Social Science Research in Law: A New Paradigm*, 43 Am. Psychologist 465 (1988).
31. Smalarz & Wells, *supra* note 1.
32. *Id.* at 21; Richard A. Wise & Martin A. Safer, *A Comparison of What U.S. Judges and Students Know and Believe About Eyewitness Testimony*, 40 J. Applied Soc. Psychol. 1400, 1400 (2010).
33. National Research Council, *supra* note 4, at 4.
34. *See* Kenneth A. Deffenbacher et al., *A Meta-Analytic Review of the Effects of High Stress on Eyewitness Memory*, 28 Law & Hum. Behav. 687 (2004); Lorraine Hope et al., *Witnesses in Action: The Effect of Physical Exertion on Recall and Recognition*, 23 Psycholo. Sci. 386 (2012); Charles A. Morgan et al., *Accuracy of Eyewitness Memory for Persons Encountered During Exposure to Highly Intense Stress*, 27 Int'l J. L. & Psychiatry 265 (2004).

line of research, one core topic at a time. Yet policy, practice, and courtroom-evidence decisions must entertain multiple lines of relevant research. For example, lineup-identification accuracy surely is affected by more than an instruction to a witness. Multiple meta-analyses each inform about a factor that will influence the quality of witness memory formed at the time of the crime event (stress, presence of a weapon, perpetrator disguise, and the like), the type of memory intrusions or loss between event and lineup and the lineup practice itself (lineup construction, lineup presentation format, lineup delivery). Just as a meta-analysis pulls together individual studies to clarify "the big picture," the expert in a scientific realm—or judge or jury—has the daunting task of weaving together the knowledge from multiple lines of research to see the broader conclusions and implications of the data.

CONCLUSION

Meta-analysis is not limited to this topic of eyewitness memory. Psychological researchers have explored an array of human behaviors relevant to law. Nor is meta-analysis unique to the behavioral sciences. Medical and epidemiological testing, educational research, and other domains use this technique. Standard criteria for published meta-analyses are available[35] as are many books and articles on the mechanics of conducting a meta-analysis.[36]

Judges can increase the efficacy of evidence-admissibility decisions by becoming familiar with a relevant research literature. Meta-analysis offers a vehicle to become quickly apprised of content in a highly specialized area. Additionally, the basics for an effective meta-analysis can provide sound expectations for expert testimony that can impart scientific-framework evidence to triers of fact.

Nancy K. Steblay, Ph.D., is Professor of Psychology at Augsburg College in Minneapolis, Minnesota. Dr. Steblay has conducted research for over 25 years and is an internationally recognized expert on eyewitness memory and identification procedures. She has published numerous peer-reviewed scientific articles, meta-analyses, and chapters on eyewitness topics. Dr. Steblay is a member of the scientist team for a national field experiment of eyewitness identification procedures sponsored by the American Judicature Society. Professor Steblay has received awards for Distinguished Contributions in Scholarship and for exemplary teaching.

35. *See*, *e.g.*, David Moher et al., *Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement*, 6 PLoS Med. e1000097 (2009). A checklist is available at http://prisma-statement.org/documents/PRISMA%202009%20checklist.pdf.
36. *See* Rosenthal, *supra* note 7.