Psychology, Public Policy, and Law
2014, Vol. 20, No. 1, 1–18

1076-8971/14/$12.00 DOI: 10.1037/law0000001

The Eyewitness Post Identification Feedback Effect 15 Years Later: Theoretical and Policy Implications

Nancy K. Steblay
Augsburg College

Gary L. Wells
Iowa State University

Amy Bradfield Douglass
Bates College

Eyewitnesses' retrospective reports of certainty, view, attention, and other judgments constitute central variables used by courts to assess the credibility of eyewitness identification evidence. Recently, important state Supreme Court decisions (e.g., *New Jersey v. Henderson*, 2011; *Oregon v. Lawson,* 2012) have relied on psychological research regarding the post-identification feedback effect to revamp their assumptions about when witness retrospective self-reports can be trusted. The post-identification feedback effect, originally demonstrated by Wells and Bradfield (1998), refers to the way in which witness self-reports are distorted by feedback to the witnesses that suggests that their identifications were accurate or mistaken. We present a meta-analysis of the post-identification effect involving 7,000 participants from the United States, Canada, Europe, and Australia. The results show that confirming feedback robustly inflates eyewitnesses' retrospective judgments across experimental manipulations and laboratory settings with large effect sizes. We describe the policy implications of the feedback effect with regard to the need for double-blind lineup procedures. Moreover, we propose that testimony-relevant witness judgments should be collected and documented, preferably with videotape, before feedback can occur. We use moderator analyses to examine the current explanation of the feedback effect and delineate new research questions that could help develop a more complete theoretical understanding of the processes giving rise to the effect.

Keywords: post-identification feedback, eyewitness memory, double-blind lineups, lineup procedure

The Oregon Supreme Court recently addressed the fundamental problem of mistaken eyewitness identification with a science-based analysis that repositioned eyewitness evidence to align with state evidence law (*Oregon v. Lawson,* 2012). The burden for eyewitness evidence reliability in Oregon is now placed squarely on the party that desires to admit such evidence at trial, namely the prosecution. This ruling requires Oregon judges to scrutinize—regardless of whether or not law enforcement used a suggestive identification procedure—whether the witness's testimony is based on personal perception and knowledge. The source of probative value in an eyewitness report is recognized to be the *original memory* of the witness, uncontaminated by outside (prejudicial) information.

Relying heavily on eyewitness identification research in psychology, the Oregon court listed post-identification feedback as one of eight system variables that can affect the reliability of eyewitness evidence: "Confirming feedback, by definition, takes place after an identification and thus does not affect the result of the identification itself. It can, however, falsely inflate witness confidence in the reports they tender regarding many of the factors commonly used by courts and jurors to gauge eyewitness reliability. As a result, the danger of confirming feedback lies in its potential to increase the *appearance* of reliability without increasing reliability itself" (*Oregon v. Lawson,* 2012, p. 21).

Prior to *Lawson,* the standard for most courts derived from a U.S. Supreme Court ruling in which eyewitness answers to questions such as "How good was your view of the culprit?" "How much attention were you paying?" and "How certain were you in your identification?" were recommended as core factors to consider in evaluating the reliability of witness identification of a suspect (*Manson v. Braithwaite*, 1977). As articulated in the *Manson* ruling, the certainty of the witness (as with view and attention to the event) is considered to be a trustworthy aspect of eyewitness evidence, a marker for reliability. The *Lawson* decision, however, 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. Emphasizing this possibility is a recent analysis of 161 DNA exoneration cases indicating that up to 57% of mistaken witnesses who testified confidently at trial had been substantially uncertain at the initial identification: 40% did not identify the defendant at first try, 21%

Nancy Steblay, Department of Psychology, Augsburg College; Gary L. Wells, Department of Psychology, Iowa State University; Amy Bradfield Douglass, Department of Psychology, Bates College.

We thank Simone Schriger for help in preparing the article.

Correspondence concerning this article should be addressed to Nancy Steblay, Department of Psychology, Augsburg College, 2211 Riverside Ave, Minneapolis, MN 55454. E-mail: steblay@augsburg.edu

admitted uncertainty, 9% said they didn't see the face (Garrett, 2011). Even so, witness confidence grew across time, culminating in convincing trial testimony being leveled against innocent individuals.

The post-identification feedback effect (hereinafter, the "feedback effect") has been conceptualized in the research literature as a system variable that has powerful impact on the retrospective judgments of an eyewitness in the immediate aftermath of a lineup decision. Fifteen years of research tell us that when a lineup administrator confirms an eyewitness's lineup identification, ("Good, you identified the suspect"), the witness is subsequently likely to render a significantly inflated retrospective report of the quality of the identification process (certainty and ease of the identification) as well as his or her viewing experience for the crime event (how good the view was, how much attention was paid to the event) and other testimony-relevant judgments (Wells & Bradfield, 1998). As recognized in *Lawson*, the post-identification feedback effect means that the *appearance* of memory reliability has been influenced by the lineup administrator. In part because courts are relying heavily on the post-identification feedback effect in remaking ground rules for eyewitness identification evidence, it is important to carefully assess the scientific reliability of the feedback effect, its consistency, strength, and robustness.

We conducted a meta-analysis of the extant feedback effect literature to provide guidance to law enforcement and courts as they further consider the implications of post-identification feedback as a system variable. Meta-analysis has been recognized by courts as a valuable aid to their decisions regarding psychological research. For example, in a recent appellate decision the New Jersey Supreme Court relied on a court-appointed Special Master who reviewed the scientific eyewitness identification literature. The Special Master, retired Superior Court judge Geoffrey Gaulkin, reported that "[a]n important and much cited subset of the literature is comprised of meta-analyses, which evaluate the methodologies and findings of multiple published reports of experiments in a given area of inquiry. [internal citations omitted] The strength of meta-analyses is dependent, of course, on the strength of the underlying studies, but because of their breadth, meta-analyses are generally regarded as offering the most reliable statements of the scientific findings" (*New Jersey v. Henderson*, 2011, Report of the Special Master, p. 12). Indeed, an earlier meta-analysis of the post-identification feedback effect was used in the *Henderson* court's appellate decision (Douglass & Steblay, 2006).

The current meta-analysis of published peer-reviewed research incorporates twice the published articles of the feedback effect tested in Douglass and Steblay (2006), some 7,000 witness-participants (2,400 were reviewed by Douglass and Steblay), and, perhaps more importantly, includes moderator variables that were not available in the literature in 2006.

This updated meta-analysis serves two functions, a policy function and a theoretical function. At the policy level we focus primarily on the idea that post-identification feedback contaminates key indicia of eyewitness identification reliability and yet there are procedures under control of the justice system (involving double-blind lineup administration and timely questioning of witnesses) than can preserve these indicia in their uncontaminated state. At the theoretical level, moderator analyses are used to examine the dominant interpretations of the feedback effect and to raise new questions about the psychological processes that give rise to the effect. We turn to the policy implications of the post-identification feedback effect first.

Policy: The Double-Blind Lineup

In the policy domain, the feedback effect holds strong implications for legal system recommendations regarding the double-blind lineup procedure. A double-blind lineup procedure is one in which the lineup administrator does not know which lineup member is the suspect and which are merely fillers. The double-blind lineup was first introduced 25 years ago as a way to prevent lineup administrators from influencing eyewitness identification decisions (Wells, 1988). The idea was based largely on the well-known experimenter-expectancy effect (e.g., Rosenthal, 1964; Rosenthal & Rubin, 1978). The double-blind lineup follows from the analogy between police conducting a lineup to test a hypothesis (that the suspect is the culprit) and researchers conducting an experiment to test a hypothesis (Wells & Luus, 1990). In both cases, the gold standard would seem to be a double-blind procedure in which all critical dependent measures are collected before the person conducting the test is aware of the hypothesis (as to which lineup member is the suspect). The American Psychology-Law Society (Division 41 of the American Psychological Association) endorsed the idea of double-blind lineups in a 1998 "white paper" (Wells et al., 1998). Many jurisdictions have now switched to double-blind lineups (e.g., the states of New Jersey, Connecticut, North Carolina, and Ohio as well as major cities and counties such as Boston, Dallas, and Santa Clara County, CA).

A concern with the possibility that nonblind administrators can influence witness identification decisions (i.e., whether and who witnesses pick from the lineup), however, is only one reason why double-blind lineups should be strongly considered among good practices. Largely forgotten as a reason for double-blind procedures is the prevention of the post-identification feedback effect, an argument that was made in the original Wells and Bradfield (1998) article.

The importance of preventing the post-identification feedback effects goes beyond the core dependent measure of eyewitness certainty. The *Lawson* decision's emphasis on the witness's original personal memory as evidence pointedly reminds us that post-identification feedback affects other aspects of witness reliability as well, such as the witness's report of attention paid to the event, quality of view, and ability to make out the features of the perpetrator's face (thereby suggesting favorable levels of "estimator" factors). A nonblind lineup administrator who influences witnesses' answers to testimony-relevant questions is influencing the evidence that triers of fact (judges and juries) use to evaluate the identification. In recent debates about double-blind lineups (see Clark, 2012; Wells et al., 2012), the post-identification feedback effect has not been a central part of the dialogue regarding the need to double-blind lineups. However, it could be argued that the feedback effect deserves a place at the table in any discussion of the blind versus nonblind policy. We explore these policy implications in the Discussion section, after reporting the meta-analytic findings.

Theory: Explaining the Post-Identification Feedback Effect

In addition to its practical value for legal policy development, the feedback effect is imbued with an element of theoretical intrigue regarding the psychological processes that give rise to the effect. This intrigue is largely the result of three central components of the effect. First, the feedback manipulation occurs *after* the identification is made, which means that feedback could not have influenced identification accuracy. Second, key dependent measures ask witnesses to make *retrospective* judgments about matters that occurred *before* the feedback. In the case of witness certainty, for instance, witnesses are asked to indicate how certain they were at the time of the identification. In the case of view and attention, witnesses are asked to indicate how good their view was and how much attention they paid at the time of witnessing. Hence, differences between feedback conditions represent memory distortions rather than reality. It should be noted that the feedback effect is more than a simple failure to remember or a superficial compensation for nonremembered details of experience; feedback distorts witness memory well beyond the boundaries of misreporting what one knew at an earlier time (i.e., the hindsight bias; see Bradfield & Wells, 2005; Fischhoff, 1975). Third, the effect is unusually broad across a wide variety of variables, covering a dozen or more dependent measures that are highly relevant to testimony.

The original interpretation of the post-identification feedback effect was that witnesses do not form online memory traces for testimony-relevant judgments such as how good or poor their view is, how much attention they are paying, how certain they are when they make their identification, and so on (Wells & Bradfield, 1998). As a result, when witnesses receive feedback suggesting they made a correct identification (even when they were mistaken), they use the feedback as a cue to infer the answers (e.g., "I made a correct identification, so I must have had a good view, paid attention, and been certain"). Variations of this interpretation have emerged in later work. For instance, instead of assuming that no memory traces for these judgments were formed, later descriptions of the process tended be more agnostic regarding the question of whether memory traces were or were not formed and instead stressed the idea that they were simply not cognitively accessible (e.g., Neuschatz et al., 2007; Wells & Bradfield, 1999). This *cue-accessibility* interpretation left open the possibility that memory traces had been formed but were no longer accessible at the time of retrieval. Common to both interpretations is the idea that the effect involves a process in which witnesses rely on the feedback as a cue to infer their view, attention, certainty, and other aspects of past experience. According to the cues account, witnesses are relegated to relying on the feedback and inferences processes largely because there is little or no accessible memory trace for making these judgments. The cue-accessibility interpretation of the feedback effect has been likened to Bem's (1972) self-perception theory of attitudes and beliefs; to the extent that internal cues are weak, people infer their attitudes and beliefs by observing their own behavior and the context.

Using this cue-accessibility framework, Wells and Bradfield (1999) implemented a "private thought" manipulation to test the idea that forcing witnesses to form an accessible "prefeedback trace" for these judgments would provide them with internal cues that would moderate the feedback effect. In what has become known as the feedback-prophylactic effect, Wells and Bradfield found that asking witnesses to privately reflect on their certainty, view, and so forth *prior* to receiving feedback serves to significantly reduce the post-identification feedback effect.

The feedback prophylactic effect not only provides support for the cue-accessibility conceptualization, but also rules out more mundane interpretations, such as witness self-presentation. A self-presentation interpretation states that witnesses who get confirming feedback are merely posturing to make themselves look good ("Yes, I was certain all along"), which they feel permitted to do once they get feedback suggesting that they made an accurate identification. But if a self-presentation motive underlies the post-identification feedback effect, the private-thought manipulation prior to feedback should not moderate the effect; witnesses could still boastfully posture and self-present because no one knew what their private thoughts were. Hence, self-presentation has not fared well as an explanation of the feedback effect.

Building on cue-accessibility, Charman, Carlucci, Vallano, and Gregory (2010) developed a more elaborate *selective cue integration framework* (SCIF) positing a three-stage process. According to Charman et al.'s SCIF account, when witnesses are asked about their viewing and identification experiences, they first assess the strength of internal cues for making these judgments (assessment stage). If internal cues are weak, witnesses then look for external cues (search stage). If external cues are found (such as in administrator feedback), then witnesses submit the external cues to a credibility check (evaluation stage). If the external cues are judged to be credible, then the external cues are used to make the judgments. This framework is modeled after the attitude-accessibility literature and is useful in explaining how manipulations that discredit the feedback (such as suggesting questionable motives of the source of the feedback or learning that the feedback was randomly generated) serve to reduce the feedback effect.

Also of theoretical interest is the question of whether witness accuracy moderates the feedback effect. The cue-accessibility conceptualization has been used to postulate that witnesses who make accurate identifications will have stronger internal cues (due to the sense of true recognition that occurred) and thereby be less influenced by external cues than are inaccurate witnesses (Bradfield, Wells, & Olson, 2002). There are now eight studies testing the feedback effect with accurate witnesses. The question of whether accuracy moderates the feedback effect is important because if accuracy does not moderate the effect, then it raises questions about the cue-accessibility conceptualization of the feedback effect.

Our goal for this theoretical analysis of the feedback effect goes beyond merely informing researchers interested in the phenomenon. Instead, we anticipate that a comprehensive picture of the theory underlying this important effect will inform future collaborations between the legal system and researchers.

Additional Considerations for the Meta-Analysis

As noted above, an early meta-analysis on the post-identification feedback effect has been part of the scientific offerings to the court. In that original analysis, Douglass and Steblay (2006) summarized the nascent *post-identification feedback effect* literature, finding a robust effect with medium to large effect sizes

for a broad set of dependent measures. Since that review, the published feedback effect literature has now more than doubled and includes studies conducted in the United States, Canada, Europe, and Australia and with both laboratory and actual witnesses to serious crimes. We anticipate that this larger literature will give more stable estimates of effect sizes for testimony-relevant judgments and permit analyses bearing on the question of how these judgments are related. Moreover, since the previous meta-analysis, more moderators have been explored in order to define the possible boundaries of the effect, and more extensive theoretical models have been developed, thereby permitting a more detailed look at how the findings agree with theories of post-identification feedback. We also expect that this meta-analysis will reveal shortcomings of the post-identification feedback literature, especially with respect to mediators and causal paths relating to the various dependent measures. Therefore, we expect our analysis to open new questions for future research, such as why the inflating power of confirming feedback is stronger than the deflating power of disconfirming feedback.

Specific to the cumulative nature of research findings, concern has been expressed in recent years about the tendency of some research effects—that appear real based on initial studies—to lessen or even disappear over time as researchers attempt to replicate and refine the original findings or to ascertain the boundaries of the effect (e.g., Ionnadis, 2008; Schooler, 2011). Moreover, narrative impressions of an empirical literature are subject to a number of cognitive biases, such as the availability bias, that can only be effectively corrected with a meta-analysis (Bushman & Wells, 2001). Given the centrality of the feedback effect in legal policy discussions and legal rulings, the reliability, robustness, and size of the feedback effect at this point, 15 years after the original publication, is of considerable interest.

Method

Sample

Our central hypothesis is that post-identification confirming feedback will produce significant witness response inflation for the broad set of dependent measures common to this research literature (Douglass & Steblay, 2006). Selection criteria for the current meta-analysis included: published experimental tests of event memory (excluding facial recognition paradigms or those testing memory for details of an event, e.g., Dixon & Memon, 2005); random assignment of participants to Feedback versus No Feedback groups; dependent measures of witness retrospective confidence and testimony-relevant variables broken out by dependent variable (not composite scores, e.g., Rodriguez & Berry, 2010); and sufficient data for calculation of effect size between Feedback and No Feedback conditions either in the published report or through contact with the author. Our primary analyses focus on 20 published articles, representing 6,200 participant-witnesses from 10 separate laboratories that met these criteria. Some articles include more than one experiment. (See Table 1 for listing of studies.) Additional published studies that failed to meet these specific criteria but that offer useful ancillary material also are discussed: for example, regarding feedback effects on real witnesses to crime (Wright & Skagerberg, 2007) and on observers of witnesses who received feedback (Douglass et al., 2010; Maclean et al., 2011). In total, 23 published articles, 11 laboratories, and 7,000 witness-participants are represented from the United States, Canada, Europe, and Australia.

Across the published articles 10 different witnessed events were used, but there was also considerable sharing of materials. Two videos were heavily used in this literature, namely the "bomber on

Table 1
Studies in the Meta-Analysis by Date of Publication: Confirming Versus No Feedback Effect Sizes (d)

		Culprit-absent lineup			Culprit-present lineup		
Study	Date	Certain	View	Attention	Certain	View	Attention
Wells & Bradfield	1998	1.04	.67	.64			
Wells & Bradfield	1999	1.05	.64	.49			
Bradfield, Wells, & Olson	2002	.65	.45	.44	.25	.09	.10
Wells, Olson, & Charman	2003	1.04	.48	.41			
Semmler, Brewer, & Wells	2004				.44		
Neuschatz, Preston et al.	2005	1.20	.62	.45			
Neuschatz, Preston et al.	2005	1.31	.47	.16			
Douglass & McQuiston-Surrett	2006	.95	.44	.36			
Douglass & McQuiston-Surrett	2006	1.14	1.02	.96	.67	1.32	.78
Lampinen et al.	2007	1.41	.48	.48	.33	.38	.40
Lampinen et al.	2007	.90	.60	.46	.68	.44	.43
Neuschatz, Lawson et al.	2007	1.14	.80	.61			
Charman & Wells	2008	.37	.39	.36			
Douglass, Brewer, & Semmler	2010	.47	.21				
Skagerberg & Wright	2009	.60	.70	.64			
Charman, Carlucci et al.	2010	1.13	.46	.60			
Charman, Carlucci et al.	2010	.87	.97	.44			
Quinlivan, Wells et al.	2010	1.39	.80	.32			
Quinlivan et al.	2012				.35	.38	−.19
Charman & Wells	2012	.90	.31	.12	.63	.61	.33
Smalarz & Wells	in press	.79	.48	.72	.37	.19	.16

the roof" video (first used by Wells et al., 2003, used in seven articles) and the "Target store security video" (first used by Wells & Bradfield, 1998, used in five articles). Eight other articles used unique events, including various theft videos, a mugging video, a bank robbery video, an airport bag-switching video, and a live event. One feedback article (Wright & Skagerberg, 2007) used actual eyewitnesses to crimes in ongoing cases in the United Kingdom. The articles did not provide enough information to determine how good or poor the witnesses' views were of the culprit. A few of the articles mentioned culprit-exposure durations (ranging from 5 s to 43 s), but information about distance, clarity, lighting, distractions (e.g., other people in the video), viewing angles (e.g., profiles, straight on), and other factors that would be needed to score videos for view were generally not reported, thereby preventing any analyses of potential moderating roles for encoding conditions across studies.

Dependent Measures

Dependent measures related to the post-identification feedback effect cover an array of testimony-relevant variables that represent three broad aspects of witness responses, as originally presented by Wells and Bradfield (1998): (a) *memory acquisition judgments,* (b) *memory retrieval judgments,* and (c) *summative judgments.* We use this tripartite categorization to organize our report on the many dependent measures that have been used to study the post-identification feedback effect.

Memory acquisition judgments refer to witnesses' retrospective assessments of the witnessing experience. Five measures fall into this category. These include witnesses' recollections of how good their view was of the culprit (*view*), how much attention they paid to the culprit during the witnessed event (*attention*), how well they could make out facial details of the culprit during the event (*facial details*), the estimated distance that the culprit was from the camera-eye view (*viewing distance*), and the amount of viewing time they had to observe the culprit (*viewing time*).

Memory retrieval judgments refer to the witnesses' retrospective assessments of the recollection and identification experience. Four measures fall into this category. These include witnesses' recollections of how certain they were at the time of identification (*certainty*), how easy they found it to make an identification (*ease*), the amount of time that it took to make an identification (*time to ID*), and how clear of an image of the culprit they had in memory (*memory clarity*).

Summative judgments refer to more global assessments that are not directly asking about acquisition or retrieval judgments. Four measures fall into this category. These include overall assessments of their willingness to testify about their identification (*willingness*), how much they would trust an identification by another person who had a similar witnessing experience (*trust in eyewitnesses*), how good of a basis they had for making their identification (*basis*), and how good their memory is for the faces of strangers (*memory for faces*).

Statistics

For analyses that combine studies, each study contributes only one effect size, thereby weighting each test equally. Cohen's *d*, the standardized mean difference between two groups, was calculated as the effect size indicator for each comparison (Cohen, 1988) and a mean *d* is used to indicate effect size across studies, following a fixed-effects model. According to Cohen, a small effect is .20, a medium effect is .50, and a large effect is .80. A meta-analytic *Z* was calculated using Rosenthal's (1991) method of combining *t*-values.

Results

Table 1 lists all the published feedback studies used in this meta-analysis according to their publication dates. For each study, effect sizes are reported for three of the 13 measures. These three measures (certainty, view, and attention) are considered particularly important because they constitute what has been dubbed "Manson factors." Manson factors are variables that were singled out by the U.S. Supreme Court in *Manson v. Braithwaite* (1977), as well as the Court's earlier ruling (*Neil v. Biggers*, 1972), as indicia of reliability for eyewitness identification cases. Sample sizes for these primary comparisons ranged from 40 to 152 (M = 86.9). Table 2 reports effect sizes collapsed across the studies for the larger set of 13 dependent measures.

Effect Sizes for Confirming Feedback to Mistaken Witnesses

Examination of post-identification feedback initially arose from concern regarding confidence inflation in *inaccurate* witnesses—those who had selected an innocent suspect from a lineup. Therefore, the forensically-relevant laboratory test involved eyewitnesses who made an erroneous identification from a culprit-absent lineup. When such witnesses are provided with confirming feedback (about the success of what is a *wrong* pick), their confidence in this decision is expected to rise significantly along with their retrospective assessment of the experience of viewing the crime event (see Table 2 and Figure 1).

Our first analysis is the specific test of a feedback effect as framed by Wells and Bradfield in 1998. Adult witness-participants view a crime event and are asked to identify the perpetrator from

Table 2
Confirming Feedback Versus No Feedback Comparison for Culprit-Absent Lineups

Dependent measure	*k*	*d*	Range min	Range max	CI 95%		*SE*
Certainty	19	.98*	.37	1.41	.84	1.12	.071
Willingness	17	.98*	.66	1.57	.85	1.11	.065
Basis	17	.90*	.49	1.59	.76	1.04	.070
Ease	17	.86*	.19	1.45	.71	1.01	.075
Memory clarity	15	.69*	.32	1.13	.59	.79	.053
Trust in eyewitness	5	.69*	.24	1.08	.36	1.02	.170
Facial details	17	.65*	.21	1.14	.52	.80	.070
View	19	.58*	.21	1.02	.48	.68	.050
Time to ID	17	.54*	.08	1.15	.40	.68	.070
Memory for faces	16	.52*	−.02	.90	.38	.64	.060
Attention	18	.48*	.12	.96	.39	.57	.047
Viewing time	7	.04	−1.00	.69	−.35	.43	.197
Viewing distance	10	.00	−.38	.27	−.06	.06	.062

* $p < .001$. (A significant difference between confirming feedback and no feedback conditions.)

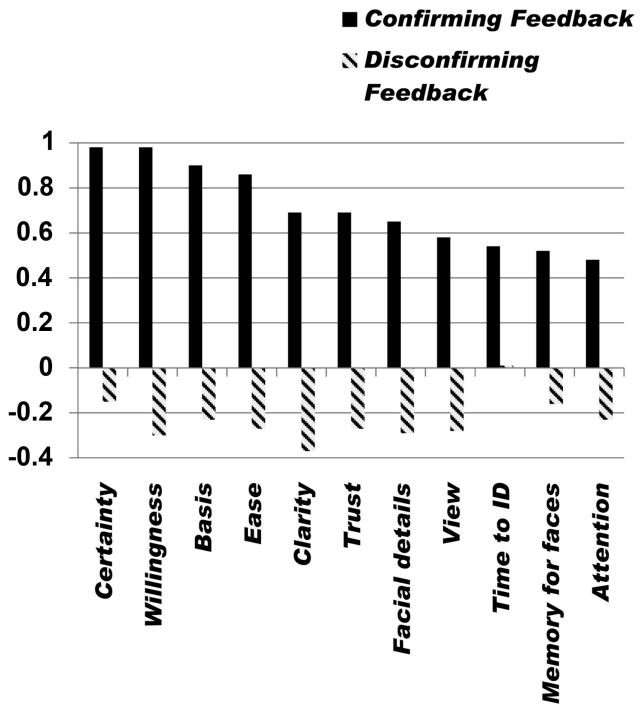

Figure 1. Effect sizes (*d*): Confirming and Disconfirming Conditions.

a culprit-absent simultaneous lineup. The witness is not provided a cautionary instruction that the culprit may not be in the lineup. Identification of any lineup member is immediately followed by the experimental manipulation via random assignment of participants to two groups: One group of witnesses is informed "Good, you identified the suspect," the other group is given no feedback. This critical feedback regarding the "correctness" of the witness' lineup pick is delivered by the lineup administrator; the witness is allowed to presume a nonblind administrator who knows which lineup member is the suspect. Each witness then completes a series of dependent measures assessing the witness's experience of the crime event and the identification procedure.

Memory acquisition judgments: Recalling the crime event. As originally conceptualized by Wells and Bradfield, (1998), memory acquisition measures include *view*, *attention*, *facial details*, *viewing distance*, and *viewing time*. As predicted, across studies witnesses who received confirming feedback reported a significantly stronger quality of witnessing experience for the crime than did witnesses who received no feedback. As shown with the effect sizes reported in Table 2, witnesses reported a better view of the event ($d = .58$, $k = 19$), greater attention paid to the culprit ($d = .48$, $k = 18$), and greater ability to make out facial details of the culprit ($d = .65$, $k = 17$), all $Zs > 8.80$, $ps < .001$. However, there was no significant difference between Confirming Feedback and No Feedback conditions on measures of viewing distance ($d = .00$, $k = 10$) and viewing time ($d = .04$, $k = 7$). This is evidence that the feedback effect is not simply a witness response bias that pervades all judgments. See Table 2 for effect sizes, effect size ranges, 95% confidence intervals around the effect sizes, and standard errors for the effect sizes for 13 dependent measures.

Memory retrieval judgments: Recalling the identification. Retrospective *certainty* is the central dependent measure in the feedback paradigm. (In this article, as in the research, the terms *certainty* and *confidence* are interchangeable.) Along with certainty, Wells and Bradfield (1998) included *ease, time to ID,* and *memory clarity*. In 19 published tests eyewitnesses who received confirming feedback expressed significantly more retrospective certainty about their positive identification decision compared to those who received no feedback, $d = .98$, $k = 19$, $Z = 18.81$, $p < .001$. Related measures aligned with this inflation of certainty. Witnesses reported a significantly greater ease ($d = .86$, $k = 17$), faster time to ID ($d = .54$, $k = 17$), and greater memory clarity ($d = .69$, $k = 17$), all $Zs > 9.80$, $ps < .001$.

Summative judgments. The impact of confirming feedback generalized to all four summative measures tested by Wells and Bradfield (1998). Compared with witnesses who received no feedback, those who received confirming feedback reported a stronger basis for their identification ($d = .90$, $k = 17$), greater trust in eyewitnesses who have a similar experience ($d = .69$, $k = 5$), a better memory for faces ($d = .52$, $k = 16$), and increased willingness to testify about their identification ($d = .98$, $k = 17$), all $Zs > 7.00$, $ps < .001$.

Effect Sizes for Disconfirming Feedback to Mistaken Witnesses

The original Wells and Bradfield study (1998) included a condition of disconfirming feedback for comparison against a No Feedback control group after all witnesses had made selections from a culprit-absent lineup. The expectation for a reversed effect—that is, for witnesses who received disconfirming evidence to be less certain—was supported ($d = -.35$), in line with the cue-accessibility notion that disconfirming feedback was used by witnesses to inform their judgments. Wells and Bradfield noted an interesting asymmetry, however, in that disconfirming feedback was less influential in lowering the confidence of witnesses than confirming feedback was in raising confidence (see Table 3 and Figure 1).

In seven studies, witnesses who received immediate disconfirming feedback after an identification from a culprit-absent lineup were compared with witnesses who received no feedback. Disconfirming feedback produced significantly lower witness ratings on 10 of 13 dependent measures, including *certainty* (see Table 3). Across measures, effect sizes are significantly smaller and more uniform ($M = .26$, $SD = .07$, range from .15 to .37) than those obtained with confirming feedback ($M = .65$, $SD = .21$, range from .41 to 1.00), $t(9) = 5.54$, $p < .001$, echoing the asymmetry noted by Wells and Bradfield (1998) in their original study.

Robustness of the Feedback Effect

Has the feedback effect size changed over time? In order to test the possibility of a "fading" effect across time, effect sizes in studies conducted as of 2006 (date of the Douglass and Steblay meta-analysis) were compared with those published after 2006. Confirming Feedback and No Feedback conditions on 10 dependent measures were included in this analysis, as each of these was tested in at least six studies for each of the two time segments. The mean effect size across the 10 measures for Confirming Feedback

Table 3
Disconfirming Feedback Versus No Feedback Comparison for Culprit-Absent Lineups

Dependent measure	*k*	*d*	Range min	Range max	CI 95%		*SE*
Certainty	7	−.15*	−.86	.44	−.48	−.18	.166
Willingness	7	−.30*	−.78	−.09	−.47	−.13	.088
Basis	7	−.23*	−.63	.17	−.48	.02	.130
Ease	7	−.27*	−.68	.00	−.45	−.09	.093
Image clarity	6	−.37*	−.67	.00	−.55	.19	.094
Trust in eyewitnesses	1	−.27*					
Facial details	7	−.29*	−.67	.00	−.47	−.11	.094
View	7	−.28*	−.63	−.09	−.44	−.12	.083
Memory for faces	6	−.16*	−.37	.03	−.30	−.02	.070
Attention	7	−.23*	−.55	.07	−.39	−.07	.082
Time to ID	7	.01	−.26	.26	−.15	.17	.080
Viewing time	3	−.16	−.35	.03	−.38	.06	.110
Viewing distance	3	.17	.00	.50	−.15	.49	.165

Note. Negative sign (−) indicates lower score in disconfirming feedback than in no feedback condition.
* $p < .05$. (A significant difference between disconfirming feedback and no feedback conditions.)

versus No Feedback groups in studies prior to 2007 was $d = .70$ ($k = 8$), the mean effect size for studies conducted after 2007 ($k = 11$) was $d = .73$. The feedback effect has been maintained across time.

Prelineup instructions. Following the lead of Wells and Bradfield (1998), most post-identification feedback studies use biased instructions implying that the culprit is in the lineup regardless of whether the culprit is present or not. That is an efficient paradigm because it leads all witnesses to make an identification. But it also raises the question of whether the post-identification feedback occurs when unbiased instructions are used (i.e., a warning that the culprit might not be present and an option to make no identification). The answer appears to be "yes." The feedback effect has been obtained when an unbiased instruction is administered (Dysart, Lawson, & Rainey, 2012; Semmler, Brewer, & Wells, 2004). Moreover, a positive lineup identification is not a necessary prerequisite to confidence inflation; that is, confirming feedback also increases the confidence of witnesses who correctly or incorrectly reject the lineup (Semmler et al., 2004).

Variations in manipulations of feedback. The standard method of manipulating feedback has been to tell the witness "Good, you identified the suspect." But some studies have used alternative ways of providing positive feedback to the witness, such as "84 of 87 witnesses agreed with you" (Semmler et al., 2004; Skagerberg & Wright, 2009) or "you've been a really great witness" (Dysart et al., 2012). The post-identification feedback effect also occurs with these alternative forms of feedback.

Generalization across participant samples. Although most researchers have tested college-age witnesses, both children (Hafstad, Memon, & Logie, 2004) and older adults (Neuschatz et al., 2005) also exhibit the feedback effect, thereby attesting to generality across age. Ear-witnesses are vulnerable to feedback effects as well, an extension of the effect across sensory modality (Quinlivan et al., 2009). Venturing even further from the standard post-identification feedback paradigm, Wright and Skagerberg (2007) tested actual witnesses to crimes in ongoing investigations in the United Kingdom. These researchers did not ask the *certainty* question because of concerns about keeping the data confidential from defense lawyers and gaining approval for the study from police. Instead, they used a proxy for certainty: "How *easy* was it for you to figure out who committed the crime?" The researchers were unable to randomly assign these real witnesses to feedback conditions, so instead asked the *ease* question either before or after the witnesses were told whether they identified the suspect or identified a filler. Witnesses told that they identified a filler reduced their ratings of *ease* of identification; witnesses told that they identified the suspect increased their ratings of *ease* of identification. The net effect of feedback was equivalent to $d = 1.25$ on witness ratings of ease. Whether *ease* is a good proxy for *certainty* is debatable. However, this study is an important demonstration of generalizability because it used a sample of actual witnesses to serious crimes (including victims and bystanders) in ongoing investigations rather than the typical participants in a laboratory study.

Other dependent measures. Although most post-identification feedback studies have restricted their measures to the standard post-identification feedback questions, there have been attempts to take other measures. Wells et al. (2003) and Wells and Bradfield (1998) explored a set of measures that tap witnesses' retrospective reports of how they made their identification decisions. Confirming feedback significantly increased witness endorsement of items related to automatic recognition: The suspect's face "just popped out at me," "I just recognized him," and "I matched the image in my head to the person in the photo." There was no significant effect for feedback on items that suggested a more deliberative process: "I compared . . .to narrow down the choices," "I first eliminated the ones definitely not him," "I based the judgment on specific facial features." Lampinen et al. (2007) also reported a significant impact of confirming feedback on witness endorsement of a "pop-out" item.

Differences in Effect Sizes Across Measures

It is clear from Table 1 that some measures show large effect sizes and other measures show much smaller effect sizes. There appears to be no pattern relating effect sizes on the measures and

whether the measure is in the acquisition, retrieval, or summative category. Hence, we explored two possible explanations for this. One possibility is that some measures are operating against ceiling effects because of their "starting points" in the control condition (no room to move up on the scale following confirming feedback). Another possibility is that the order in which the questions are asked influences the size of the effect obtained on a given measure.

Differential ceiling or floor effects across measures? Some measures might have weaker effects from confirming feedback because their control condition (No Feedback) means are already high on the scale, which would leave little room for additional boosts from confirming feedback. To explore this possibility, mean scores in the control condition were transformed to a common 1–10 scale (midpoint 5.5) for each dependent measure in each study and then averaged across studies, weighted by study sample size. Consistent with this, we note that the control condition mean for the *willingness* measure, which shows the largest effect size, is also the lowest mean in the control condition (3.4). But even the highest control condition means were only 5.8 (*memory, ease*, and *view*) on the 10-point scale, which is barely beyond the midpoint of the scale (5.5). Hence, there seems to be plenty of room for upward movement on all the measures.

A better test of whether the starting point (control condition means on the measures) can account for different effect sizes across measures is to examine whether the measures that show larger versus smaller effect sizes in the upward direction are the same measures that show larger and smaller effect sizes in the downward direction. To test this, we correlated the absolute value of effect sizes for confirming feedback with absolute value of effect sizes for disconfirming feedback across the measures. This is a critical test because the *direction* of change for disconfirming feedback is opposite to the direction of change for confirming feedback and yet the control means are the same. In other words, if the effect of confirming feedback on a low control condition mean (e.g., willingness at a mere 3.4 on the scale) is due to there being more room to "move up" on the scale than the other measures, then there should be less room to move down on the scale than the other measures. Hence, if ceiling or floor effects explain the variation in effect sizes, the correlation between the absolute values of the effect sizes in the confirming and disconfirming conditions should be negative. In fact, however, the correlation is positive, $r(13) = .58$, $p = .04$. In other words, the measures that are most malleable for confirming feedback (which increase relative to the control) are the same as those that are more malleable for disconfirming feedback (which decrease on the scale). The *willingness* measure is a salient example. *Willingness* is the lowest control group mean (3.4 on the 10-point scale) and rises more than any other measure in response to confirming feedback. But it is also the one that drops the most in response to disconfirming feedback. Hence, ceiling and floor effects fail to explain why some measures are more influenced by feedback than are other measures.

Order effects for the measures? Another possible source of variation in effect sizes across measures concerns the order in which the questions are asked. The strong tendency of researchers has been to mimic the measures that were originally used by Wells and Bradfield (1998), including the order in which the questions are asked. Accordingly, the order of the measures is confounded with the type of measure. One can imagine different hypotheses about the order of the items as they relate to effect size. One possibility is that the first measures will show the strongest effects because of their closer temporal proximity to the feedback manipulation and the effect might then wear off on later measures. Or, one might predict the opposite, namely that there is a "warm-up" effect of some sort so that the early measures show weaker effects and the later measures show stronger effects. In fact, however, neither of these hypotheses seems to fit the data. In almost all studies (94%), *certainty* is the first item asked of witnesses, *basis* is placed in the middle of the questionnaire (Position 5, 6, or 7), and *willingness* is at or near the end. But these three measures are the strongest effects. In order to assess this more precisely, we coded the position of each of the 13 dependent measures on the witness questionnaire for each study and calculated a correlation between question order and effect size across measures. This correlation was not significant, $r(11) = .09$, $p = .78$. Thus, order of the measures does not seem to provide a satisfactory explanation for variability in effect sizes.

Credibility-Threshold Estimates: Communicating the Forensic Importance

It can be argued that eyewitnesses who will "survive" to ultimately testify against a defendant at trial have to pass a threshold of credibility. For example, a witness who expresses little certainty or reports a poor view after making an identification is not likely to impress a prosecutor enough to produce an indictment, be perceived as credible enough to survive a pretrial hearing on admissibility, or convince a jury that his or her identification should be trusted. In light of this, Wells and Bradfield (1998, Experiment 2) developed a unique measure that expresses the impact of confirming post-identification feedback on the probability that a mistaken identification witness will meet the threshold as a credible witness. Wells and Bradfield calculated the percentage of mistaken witnesses who responded at the high end of their 7-point scale (with a 6 or 7) among witnesses who did not get confirming feedback versus those who received confirming feedback. The results were startling; 50% of mistaken witnesses who received confirming feedback rated their certainty at the high end of this scale, compared with only 15% of witnesses who received disconfirming feedback. There was no control condition in Experiment 2 so the only comparison possible was between disconfirming and confirming feedback conditions.

In this meta-analysis we did not have the raw data to exactly duplicate the method used by Wells and Bradfield (1998) for all of these studies. However, we can estimate these credibility-threshold percentages for the data set as a whole using the means, effect sizes, and standard deviations of the individual studies and relying on the assumption that the data are normally distributed. The normal distribution assumption permits us to estimate the percentage of witnesses in both the control and confirming feedback conditions who fall in the upper tail of the distribution corresponding to scores of 8 or more on the 10-point scale for each measure.

Figure 2 shows the results of our analysis of the percentage of witnesses who pass the credibility threshold of 8 or greater for each of the significant measures. On certainty, for example, only 6% of the mistaken witnesses in the control condition met or exceeded the credibility threshold whereas 29% of witnesses in the confirming condition passed the threshold. Notice as well that very

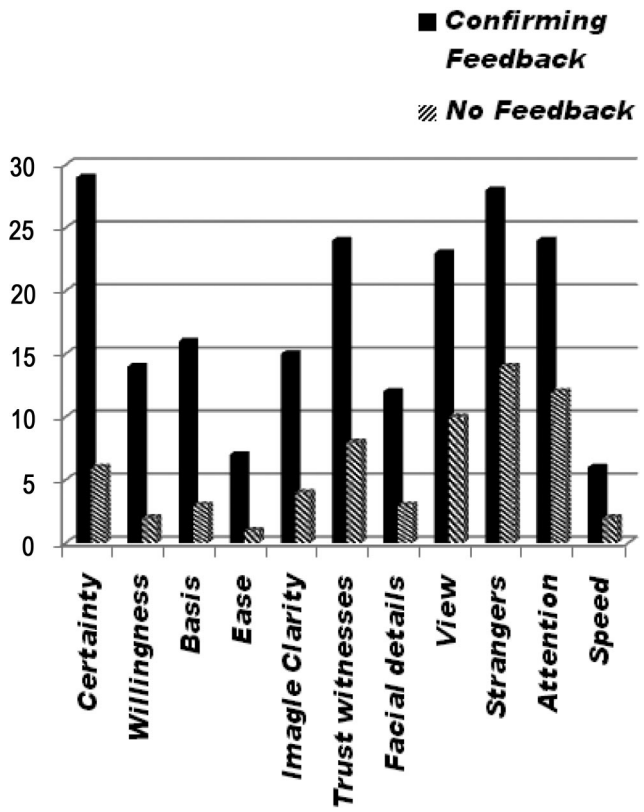

Figure 2. Estimated percentage of mistaken witnesses responding at 8 or above (on a 10-point scale).

few (2%) of mistaken witnesses passed the willingness-to-testify threshold in the absence of confirming feedback, but 14% of mistaken witnesses passed that threshold if they received confirming feedback. This pattern repeats across the measures.

This measure of the probability of passing a credibility threshold is a forensically relevant way to think about the effect size resulting from a single comment from a lineup administrator. In fact, we believe that the credibility-threshold metric is a better way to communicate the impact of post-identification feedback to policymakers (e.g., "The percentage of mistaken witnesses who will display high certainty rises from a mere 6% to 29% with feedback) than traditional effect-size metrics (e.g., "Feedback increases the certainty of mistaken eyewitnesses by .98 standard deviations").

Mitigating the Post-Identification Feedback Effect

There have now been 10 attempts to mitigate the post-identification feedback effect using manipulations that are either prophylactic (prior to feedback) or remedial (after feedback).

Successful mitigation is indicated if the application of the intervention manages to significantly reduce the impact of the feedback (a comparison of the intervention vs. not between witnesses who all received confirming feedback). Full mitigation is indicated if the manipulation manages to fully eliminate the difference between the Feedback and No Feedback (control) conditions. For ease of reporting, we focus below on the key dependent measure of witness identification certainty. The patterns discussed are most evident with the certainty measure but also appear throughout additional dependent measures.

Prophylactic moderation attempts. The confidence prophylactic is employed by asking eyewitnesses, after their identification but *prior* to their receiving feedback, to consider their confidence in the identification (as in Wells & Bradfield, 1998) or, more fully, to think privately about how clearly they could see the gunman's face in the video, how much they focused on the gunman's face, how easy it was for them to select someone from the photos, how good they are at remembering faces, and how sure they were that they identified the right person in the photo-spread (as in Wells & Bradfield, 1999). The objective is to produce a retrievable memory trace that will "inoculate" the witness against the influence of the subsequent feedback. The theory behind this manipulation is closely tied to the cue-accessibility conceptualization of the feedback effect.

Four studies speak to the prophylactic hypothesis. Wells and Bradfield (1998, 1999) examined eyewitnesses using confidence prophylactics described above. Quinlivan et al. (2009) required ear-witnesses to rate their certainty on a Likert-type scale prior to feedback. Although the ear-witnesses produced consistently larger effect sizes across measures compared with the eyewitnesses (perhaps because audio cues are weaker than visual cues), the pattern of outcomes is quite similar among these three studies that compared the impact of a confidence prophylactic across witnesses who all received confirming feedback. For both eyewitnesses and ear-witnesses, a confidence prophylactic as intervention reduced the impact of the feedback. Specifically, witnesses who were asked to consider their experience before they heard confirming feedback showed significantly smaller feedback effects on retrospective certainty compared with witnesses who did not receive the prophylactic instruction before they heard confirming feedback, $d = .52$. However, even with the confidence prophylactic applied there was still significant confidence inflation from confirming feedback when compared with witnesses who never heard the feedback ($d = .59$). Thus, in these three studies the confidence prophylactic moderated but did not fully eliminate the feedback effect.

The fourth study (Neuschatz et al., 2007) was able to reduce confidence inflation to the level of a No Feedback group, $d = .12$. But Neuschatz et al. also report one of two tests that expose additional limitations for the confidence prophylactic. When measures were delayed for 1 week, the feedback effect emerged at that time despite the prophylactic administered earlier. Quinlivan et al. (2009) also found that the impact of an otherwise successful confidence prophylactic disappeared for ear-witnesses on all dependent measures when the measures were delayed for 1 week. Presumably, this rebound effect occurs because witnesses' memories for their prefeedback thoughts (the prophylactic) become less accessible over time whereas their memory for the feedback remains salient and largely accessible.

Remedial moderation attempts. An alternative strategy to correct for the feedback effect is through a manipulation following the feedback that attempts to undermine the usefulness or credibility of that feedback. Three studies (Charman et al., 2010; Lampinen et al., 2007; Quinlivan et al., 2010) have followed confirming feedback with a correction to the witness that the feedback was in some manner a mistake. This "mistake" interven-

tion significantly reduced witness confidence inflation compared to a group without the intervention, $d = .85$. Yet, significant differences were sustained between Confirming Feedback/Intervention and No Feedback groups; that is, the intervention did not erase the impact of the confirming feedback ($d = .38$).

Credibility of the feedback can be undermined by interjecting suspicion about the source and intent of the feedback after the feedback is delivered. In two studies (Neuschatz et al., 2007; Quinlivan et al., 2010), the credibility of the feedback was placed in doubt by later informing the witness that the feedback source did not actually know which lineup member was the correct person and had a manipulative intent (". . . she has no way of knowing if you picked the correct person . . . She is trying to influence the results." Neuschatz et al., 2007, p. 235). This manipulation significantly reduced response inflation (Confirming Feedback/Intervention vs. Confirming Feedback, $d = .77$) within an immediate timeframe and even when received with a 1-week delay after the feedback. But, the effect remained ($d = .46$) when the Confirming Feedback/Intervention group was compared with the No Feedback group, indicating that the suspicion manipulation did not completely eliminate the impact of confirming feedback.

Blind lineup administration. The prophylactic manipulations and remedial manipulations described above produced similar results: a reduction but not elimination of the feedback effect. Two additional attempts to mitigate the feedback effect can be noted, one successful and one not, but both important to these considerations. Lampinen et al. (2007) explored a forensically realistic instruction delivered after the feedback: that the witness should ignore the feedback and base answers on his or her own best recollection. This intervention did not reduce the impact of confirming feedback.

On the other hand, a more subtle means of undermining feedback information is to lead witnesses to believe up-front that the lineup administrator does not know who the suspect is in the lineup. Dysart, Lawson, and Rainey (2012) found that when witnesses believed that the lineup administrator knew who the suspect was, the typical significant response inflation followed positive feedback (*certainty, d = .70*). Comparatively, witnesses who received positive feedback from a presumed-blind lineup administrator did not inflate any measures (*certainty*, $d = .05$). Thus, the perceived-blind administrator's feedback was seemingly disregarded by witnesses. It should be noted that this study also employed a more subtle feedback of "you've been a good witness" for all witnesses, the latter as a means to provide realistic feedback consistent between blind and nonblind lineup administrators and relevant to both accurate and inaccurate witnesses.

Psychological Processes: Confirming Feedback

As discussed above, predominant theoretical explanations for the feedback effect rest on the notion that witnesses rarely possess a memory trace or at least have no immediate access to memory for testimony-relevant aspects of their viewing and identification experiences. *Witness reports of feedback influence* are consistent with this cue-accessibility hypothesis, in that witnesses who report that they were not influenced by the feedback show response inflation that is just as strong as witnesses who report they were probably affected by the feedback (Wells & Bradfield, 1999). Furthermore, *witness reports of current confidence* are indistinguishable from their reports of retrospective confidence, providing additional support for the notion that witnesses must rely heavily on current information to infer past experiences (Semmler et al., 2004).

The cue accessibility conceptualization has led researchers to posit that witnesses who have stronger memory of the experience (internal cues) may be less vulnerable to the feedback effect. The witness's memory experience has been operationalized in a number of ways as a means to tap or to strengthen internal memory. For example, the prior-thought manipulation is a means to facilitate a prefeedback memory trace that will reduce reliance on external cues. Other means of manipulating the strength of the internal memory are described below, including testing accurate witnesses, facilitating intentional learning, and manipulating delay.

Accurate eyewitnesses. Witnesses who correctly identify the culprit from the lineup are likely to experience the identification process differently from those who make an identification from a culprit-absent lineup. More precisely, an experience of ecphory (match between memory and a lineup photo) at the time of identification should make the accurate witness less inclined to be influenced by confirming post identification feedback. Eight tests have examined witnesses who accurately identified the culprit from a *culprit-present* lineup and who received *immediate* Confirming Feedback versus No Feedback (see Table 4).

Two questions are germane regarding these *identifiers*. First, a practical question regarding generalizability of the feedback effect: Does the feedback effect occur with accurate witnesses? Among accurate eyewitnesses, those who received confirming feedback produced significant differences compared with those who received no feedback, on 10 of 13 dependent measures. For example, medium effect sizes were obtained for *certainty*, *basis for the ID*, *willingness to testify* and *view,* ranging from $d = .47$ to .54; *attention* produced a significant but smaller $d = .29$. Therefore, yes, a significant feedback effect is present among accurate as well as inaccurate witnesses. As noted by a reviewer, this should not be surprising in that memory strength and internal cues vary even among accurate witnesses. Second, a theoretical question: Is the

Table 4
Confirming Feedback Versus No Feedback Comparison for Culprit-Present Lineups

Dependent measure	*k*	*d*	Range min	Range max	CI 95%		*SE*
Certainty	8	.47*	.25	.68	.35	.59	.060
Willingness	7	.54*	.26	1.04	.32	.76	.111
Basis	7	.54*	.30	.85	.36	.72	.092
Ease	7	.45*	.15	.79	.30	.69	.074
Memory clarity	6	.40*	.00	.78	.14	.66	.135
Trust in eyewitness	1	.47*	.47	.47			
Facial details	7	.44*	−.01	.88	.20	.68	.123
View	7	.49*	.09	1.32	.19	.79	.153
Memory for faces	7	.18*	−.29	.49	−.02	.38	.103
Attention	7	.29*	−.19	.78	.06	.52	.115
Time to ID	7	.10	−.04	.38	−.05	.25	.077
Viewing time	2	.16	−.08	.40	−.31	.63	.240
Viewing distance	3	.20	−.30	.50	−.30	.70	.253

* $p < .05$. (A significant difference between confirming feedback and no feedback conditions.)

feedback effect consistently smaller for accurate versus inaccurate witnesses on these 10 measures, as would be predicted from the cue accessibility hypothesis? Studies that offer a comparison of inaccurate to accurate eyewitnesses revealed a mean effect size for accurate witnesses (across nine reported dependent measures) that is significantly smaller ($M = .43$, $SD = .11$) than for inaccurate witnesses ($M = .70$, $SD = .19$), $t(16) = 3.75$, $p = .002$.

In an additional examination of accurate witnesses, Quinlivan et al. (2012) found a significant impact of feedback effect after a 1-week delay between feedback and measures. Accurate witnesses presumably relied more on the feedback after a retention interval, affecting their ratings, for example, on *certainty* ($d = .59$) and *basis for the ID* ($d = .50$) compared with accurate witnesses who were measured immediately after the feedback (ds = .35 and .38, respectively).

Intentional-learning: One-at-a-time lineup presentation. Douglass and McQuiston-Surrett (2006) tested two variations based on a cue-accessibility framework. The manipulations attempted to prompt witnesses to better focus on the experimental tasks, as a means to enhance memory traces and reduce reliance on the subsequent administrator feedback. However, neither an intentional learning instruction prior to the crime event ("you will be asked to make an ID") nor the viewing of lineup members one-at-a-time to "spontaneously generate memorial traces for testimony-relevant judgments" (p. 997) were able to inoculate witnesses against the feedback effect.

Delay/retention interval. Most studies have involved a sequence of events in an immediate timeframe: crime event, lineup identification, dependent measures. Alternatively, some manipulations explore the impact of memory traces over time with the expectation that witnesses with poor memory (longer interval) will rely more heavily on the feedback. The results of this work show that the feedback effect is maintained but not significantly increased over these delay intervals. For example, measures of certainty, view, and attention are reported as follows across delays: 24 hr between crime and ID, ds = 1.05, .32, .77, respectively (Hafstad, Memon, & Logie, 2004); 48 hr between ID and feedback/measures, ds = 1.15, .63, .56 (Wells et al., 2003); 48 hr between ID/feedback and measures, ds = 1.28, .71, .51 (Wells et al., 2003); 1 week between feedback and measures, ds = .97, 1.22, .95 (Neuschatz et al., 2007). Accordingly, although these studies speak to the generalizability of the post-identification feedback effect over various delay intervals, they do not provide any additional support for the cue-accessibility conceptualization. There may be two reasons for this. One possibility is that the delays are simply too short and little forgetting has occurred. But, another possibility is that there is no opportunity for the internal cues to have decayed with time because there was no significant memory trace for these internal cues in the first place. In fact, this latter interpretation is consistent with the one exception about the impact of delay on the feedback effect, namely that accurate witnesses show an increased vulnerability to feedback effects with delay (Quinlivan et al., 2012). Recall that the presumption is that accurate witnesses have access to an internal cue (the ecphoric experience) that inaccurate witnesses lack, at least accurate witnesses are presumed to have this access in the immediate aftermath of the witnessed event.

Discussion

Fifteen years of empirical research on the post-identification feedback effect has revealed a robust and large impact of a lineup administrator's comments upon a witness's retrospective memory report regarding the crime event and the lineup identification procedure. Confirming feedback significantly inflates eyewitness reports on an array of testimony-relevant measures, including attention to and view of the crime event, ease and speed of identification, and certainty of the identification decision. This meta-analysis has firmly established that the feedback effect has not diminished in study findings since the original 2006 meta-analysis (Douglass & Steblay, 2006), even as researchers have attempted to determine the limits of the effect. Indeed, the effect holds across laboratories and variations in experimental method, lineup procedure, and witness samples (including real witnesses to crime). The effect is found for both accurate and inaccurate witnesses, and even for witnesses who do not choose from the lineup. The passage of time between identification and feedback, or between feedback and measures, does not weaken the effect, at least within the time intervals tested. And, the post-identification feedback carries over to witness testimony as evidenced by observers being more likely to believe oral testimony from witnesses who have received confirming feedback (Douglass, Neuschatz, Imrich, & Wilkinson, 2010; Maclean, Brimacombe, Allison, Dahl, & Kadlec, 2011; Smalarz & Wells, in press).

Importantly, confirming feedback is especially powerful when witnesses identify an innocent person from a culprit-absent lineup. Moreover, this meta-analysis has established that although attempts to prevent the post-identification feedback effect (e.g., prophylactic treatments) and attempts to undo the effect (e.g., remedial treatments) moderate the magnitude of the effect, they do not eliminate the effect. These stable patterns prompt new policy and theoretical considerations.

Policy: Current and Future

We began this article with the observation that U.S. courts are taking careful notice of the post-identification feedback effect because of its implications for confounding the assessment of eyewitness identification evidence. The standard approach to evaluating the reliability of eyewitness identification evidence was set forth by the U.S. Supreme Court in 1977 (*Manson v. Braithwaite*, 1977) and that case set the model for individual states. Central to the *Manson* ruling was the assertion that the certainty of the witness at the time of identification, the witness's attention paid at the time of witnessing, and the witness's opportunity to view the culprit during the crime are key factors for assessing reliability. Under the right conditions, witness self-reports of these factors are likely to be useful cues to the reliability of an identification. The post-identification feedback effect, however, threatens this central premise of *Manson*, because confirming feedback strongly inflates witness self-reports, leading mistaken eyewitnesses to report misleadingly high levels of certainty, view, and attention (see Table 1). A highly detailed treatment of the U.S. Supreme Court's *Manson* ruling and how the post-identification feedback effect undermines *Manson*-like tests can be found elsewhere (Wells & Quinlivan, 2009b). A major purpose of the current article was to assess the reliability, robustness, and magnitude of the post-identification feedback effect to make sure that courts are relying

on the current state of knowledge concerning the effect. The results of this meta-analysis support the idea that courts should treat this post-identification feedback problem very seriously.

The problem confronting courts regarding the post-identification feedback effect is complex, and solutions at the court level are not clear. In general, once a witness has been contaminated by post-identification feedback, an attempt to assess the reliability of eyewitness identification by asking questions of that witness about the feedback appears to be an ineffective strategy. That is, witnesses who say they were not influenced by feedback (which is the majority) are influenced as much as witnesses who report that they were influenced (Wells & Bradfield, 1998). Furthermore, it is not always possible to discover whether post-identification feedback has occurred; the discovery of contaminating post-identification feedback requires that the witness or the lineup administrator can recall the feedback and faithfully report it or that the identification procedure is recorded. Note as well that the problem is not simply that the courts have formally endorsed the use of certainty, view, and attention as indicators of reliability. Even without court endorsement, people (and therefore juries) naturally use such reports as indicators of whether or not to believe the eyewitness, and they may allow one *Manson* criterion to affect their ratings of eyewitnesses on other criteria for which they have no information (Bradfield & Wells, 2000). Simply put, observers of witness testimony are much more likely to believe mistaken eyewitness identification testimony contaminated by confirming feedback than mistaken testimony that was not contaminated (Maclean et al., 2011; Smalarz & Wells, in press). This effect occurs even when observers are warned against basing their judgments of the witness on feedback delivered by the lineup administrator (Douglass et al., 2010). Thus, a better solution for the justice system is to measure the relevant variables (witness statements of certainty, view, etc.) *before* they can be contaminated and make these statements available to all who are charged with evaluating the reliability of the identification.

The post-identification feedback effect is highly relevant to policy not only at the level of the courts, but also at the level of police practice. Police jurisdictions across the U.S. have been considering their policies and procedures regarding how to collect and preserve eyewitness identification evidence so as to maximize its reliability. A central consideration has been whether to adopt double-blind lineup procedure, a recommendation based largely on the idea that a nonblind lineup administrator can inadvertently influence the choice that the witness makes from the lineup (Wells, 1988; Wells et al., 1998). Although the post-identification feedback effect has been cited as yet another reason for using double-blind lineup procedures (e.g., Wells & Bradfield, 1998), the feedback effect tends to be absent in most policy discussions regarding double-blind lineup procedures.

We believe that the post-identification feedback effect should take a more central place in policy discussions regarding double-blind lineup procedures. In particular, we endorse the idea that a blind lineup administrator should be the one who secures a statement of certainty from the witness at the time of any identification so as to preserve a record of the witness's certainty before any later feedback can occur. Double-blind lineup administrators would naturally avoid reinforcing the witness (e.g., "Good job, Mrs. Jones, that is the guy!") because they might be reinforcing the choice of a filler.[1] A second procedural consideration is to explicitly tell witnesses that the lineup administrator does not know which person might be the suspect (i.e., letting the witness know that the procedure is double-blind). Dysart et al. (2012) found that somewhat ambiguous feedback ("you've been a great witness") inflated witness's retrospective certainty but this effect was nullified by telling witnesses that the lineup administrator did not know which person was the suspect.

We contend that a *non*-blind lineup administrator who is merely instructed to not influence the witness or provide any feedback is an inadequate safeguard against influence. An important study by Garrioch and Brimacombe (2001) illustrates our point. In their experiment, lineup administrators were randomly assigned to believe (erroneously) that the culprit in a lineup was either Number 3 or Number 5. Each lineup member was then given a specific, nonbiased protocol to follow in administering a lineup to witnesses that included obtaining a certainty statement from the witnesses. When the witnesses identified a person who the lineup administrator was led to believe was the suspect, witnesses expressed higher certainty than when they identified that same person but the lineup administrator was led to believe that it was not the suspect. This contamination of witness certainty occurred despite the lineup administrators being given an unbiased protocol to follow and despite the fact that 100% of the lineup administrators and 95% of witnesses denied that the lineup administrator gave any feedback. The adoption of double-blind lineup procedures in which the blind administrator secures the relevant statements at the time of identification is a better way to deal with the post-identification feedback problem.

The recommendation to secure a certainty statement by a blind lineup administrator at the time of any identification has been advocated for 20 years (Wells, 1993) and is endorsed in the American Psychology-Law Society "white paper" on lineups (Wells et al., 1998). However, we believe that the current evidence supports an even stronger recommendation, namely that the blind lineup administrator secure not only a *certainty* statement but also statements from the witness regarding *view*, *attention*, *willingness to testify*, and the *basis for the identification*. All of these statements are highly relevant for assessing the reliability of the identification but are severely compromised by external suggestion that derives from the inevitable feedback that occurs later. That is, lineup administrators are not the only source of feedback; feedback can occur any time between the identification and subsequent testimony. In fact, merely learning that the identified person was charged with the crime is a form of post-identification feedback to a witness. There is a need for prosecutors, judges, and juries to establish whether disparity exists between a witness's report at the time of identification and a later report after witness exposure to reinforcement and external influences. Statements at the time of identification would provide this information, and of course, be discoverable and made available to the defense before trial.

[1] The identification of fillers constitute about one third of all identifications made by witnesses in actual cases according to estimates from archival studies (Behrman & Davey, 2001; Behrman, & Richards, 2005; Horry, Halford, Brewer, Milne, & Bull, in press; Horry, Memon, Wright, & Milne, 2012; Klobuchar, Steblay, & Caligiuri, 2006; Memon, Havard, Clifford, Gabbert, & Watt, 2011; Valentine, Pickering, & Darling, 2003; Wells, Steblay, & Dysart, 2013; Wright & McDaid, 1996; Wright & Skagerberg, 2007).

We further propose that policymakers conceptualize post-identification feedback as a form of evidence contamination. Psychologists have long recommended that witness memory be treated as trace evidence, in the same way that blood or crime scene footprints provide evidence of a crime (Wells, 1995). In the case of eyewitness identification, evidence is not simply *who* the witness identified from a lineup; it is also what the witness reports about identification certainty. But the problem runs deeper. A lineup administrator who confirms a witness's lineup identification ("good, you picked the suspect") has influenced eyewitness evidence regarding the crime scene, moving the witness from a report of personal memory to a version tainted by external information. In effect, the witness's report of "estimator variables" (e.g., ability to see the features of the culprit at the time of the crime) that are usually considered out of the control of law enforcement are in fact being pushed around by external feedback. We also note that our results underscore that a rejection of the memory-as-trace-evidence argument, as recently occurred in an appellate decision in New Jersey (*New Jersey v. Henderson*, 2011, p. 122), in effect compromises the legal system's ability to fully benefit from psychological research on eyewitnesses (see the Special Master's Report, 2011, p. 81 for an alternative statement on the memory-as-trace-evidence argument).

Feedback affects witness prospective judgments as well. This meta-analysis reveals willingness-to-testify as a perhaps underappreciated measure. The feedback effect on *willingness* is among the strongest of all the effects, with a mean effect size of a full standard deviation. The willingness of the average witness who makes a mistaken identification to testify against that person is dramatically inflated by a simple confirming comment from a lineup administrator. This is an important and perhaps shocking outcome. A witness's increased willingness to testify may move an investigation forward in the direction of the identified suspect. Witnesses who indicate high levels of willingness to testify are those likely to be called to the stand, to readily make themselves available to the prosecutors, and to show little or no reluctance to criminally implicate the defendant at trial. In short, any presumption of the legal system that the willingness of an eyewitness to testify against a criminal defendant is a product of the trustworthiness of the witness's memory is undermined by feedback. Moreover, in other domains of research, such as research on risky behavior, willingness to engage in a behavior is a better predictor of actual behavior than are other measures, such as intent to engage in the behavior (e.g., Gerrard, Gibbons, Houlihan, Stock, & Pomery, 2008). This leads us to strongly recommend that witnesses be explicitly asked at the time of any identification (by a blind lineup administrator before they could receive any feedback) how willing they would be to testify that the identified person is in fact the culprit. This should be a matter of record, discoverable by the defense, and usable at any later trial or hearing.

Theory: Unanswered Questions and Future Directions

The results of this meta-analysis are largely consistent with the cue-accessibility conceptualization of feedback effects. For example, having witnesses give thought to the relevant judgments prior to feedback serves to moderate the feedback effect, presumably because it creates a prefeedback memory trace that can serve as an accessible cue for answering the retrospective questions. In recent years, there have been two refinements to the cue-accessibility framework and both have some support in the results of this meta-analysis. First was the postulation that a clear ecphoric experience (strong recognition experience at the time of identification) can be an internal cue that is used to infer various judgments, which should make the witness less influenced by the external cue of feedback. Little was known about this at the time of the previous meta-analysis, but the current analysis of accurate and inaccurate witnesses is consistent with that argument. Specifically, the effect of confirming feedback is consistently stronger across measures for inaccurate identifications (a weak ecphoric experience) than for accurate identifications (a stronger ecphoric experience). This pattern is apparent from comparing effect sizes in Table 2 with those in Table 4. A second refinement to the cue-accessibility conceptualization was the *selective cue integration framework* posited by Charman et al. (2010). This framework, discussed earlier, is supported in the consistent evidence that undermining of feedback credibility (via mistake or suspicion manipulations) moderates the feedback effect.

The dependent measures. Although the meta-analytic results are broadly consistent with theorizing about the post-identification feedback effect, there remain many unanswered questions. First, the cues conceptualization is generally silent with regard to what kinds of judgments will and will not be affected by post-identification feedback. And, of the 13 commonly measured judgments, two are not affected by post-identification feedback, namely estimates of how long the culprit was in view and how far away the culprit was during the witnessed event. In fact, the lack of an effect on these measures was noted even in the first post-identification feedback study (Wells & Bradfield, 1998) and the overall effect size for these two measures in the meta-analysis is functionally zero.

The cue-accessibility framework would supposedly explain the absence of an effect on these measures by postulating that witnesses have an accessible memory trace (internal cue) that witnesses can rely on for these two judgments and therefore are unaffected by the external cue of feedback. But, in the absence of some type of independent measure of accessibility, this is circular reasoning (these measures were unaffected, so there must have been accessible internal cues; other measures were affected so there must not have been an accessible internal cues). Even if the explanation is correct, the cue-accessibility conceptualization (as currently construed) fails to be able to predict a priori what kinds of judgments will and will not be affected.

One avenue for investigation of this issue is to further examine measures that are not affected by feedback. At one level it seems a bit surprising that the viewing-time and distance measures are not affected given that the general question about *view* ("How good was the view you had of the culprit?") is consistently affected ($d = .58$) as is the question about *time to ID* ("How long did it take you to make an ID;" $d = .54$). One possible explanation is that the distance-estimate measure and the viewing-time measure are the only two of the 13 measures that have primarily involved an "objective" scale for responses. The distance measure asks for an estimate in feet and the viewing-time measure asks for time in seconds. Contrast such scales with the Likert-type characteristics of the other measures. For example, time to ID is anchored by endpoints of *a little time* to *a long time*, view is anchored by *very poor* to *very good*, attention is anchored by *none* to *total attention*.

But a study by Douglass et al. (2010) directly tested this possibility by using subjective scales for distance and viewing time (a 6-point scale from *a short distance* or *a short time* to a *long distance* or *a long time*). This format change appeared not to matter; these measures still were unaffected by feedback.

One reviewer noted that people are not good at estimating either distance or time, and we agree. But it is probably the case that people are not good at estimating any of these other variables either. Moreover, the viewing-time question is not the only question about time. Witnesses are routinely asked to estimate how long it took them to make an identification (the time to ID question), which shows moderately strong effects of feedback (d = .54). Hence, at this point our theoretical understanding of what measures will and will not be affected by feedback remains incomplete.

A related question that neither the cue-accessibility conceptualization nor the empirical work has addressed is the precise nature of the causal connections between feedback and the various judgments. Cue-accessibility presumes that each judgment is an independent inference from the available cues. However, there are many possible causal chains and mediated relations among the measures. For example, feedback may directly affect all three categories of questions (memory acquisition, memory retrieval, and summative judgments), a premise suggested by the significant effects that occur within each of these categories. Alternatively, retrospective judgments may mediate summative judgment reports (very likely in the case of *testify*, although there may be a direct effect of feedback as well). Or, in a more linear cumulative chain, witnesses may infer their certainty from the feedback and then infer other judgments (e.g., view, attention, willingness, clarity) from their certainty, or alternatively, the reverse may occur. The causal chain by which these measures affect one another has not yet been explored, and these meta-analytic data do not offer definitive answers. The cue-accessibility conceptualization presumes that the witness repeats the same inference process over and over again for each judgment but the feedback might affect a subset of judgments and these judgments in turn affect the other judgments. No one has been doing this type of work with the post-identification feedback effect, so we cannot speculate on this part of the process. Although we noted that the order of questions was not systematically related to the magnitude of the effect, it is important to note that the studies routinely use the identical or nearly identical order of measures that were used originally by Wells and Bradfield (1998). A systematic manipulation to the order of the judgment measures might be a good starting point for testing the idea that each judgment is an independent repeating of the same process rather than a process by which some judgments mediate other judgments.

Witness self-persuasion: Another aspect of the feedback effect. More broadly construed, the cues accessibility framework fails to explain the relative ease by which a simple and seemingly helpful comment by the administrator can quickly infiltrate many aspects of witness recollection. A related weakness is an inability to explain why prophylactic or remedial steps cannot fully eliminate the effect once it takes hold. An explanation for the spread of feedback effects across multiple measures as well as the difficulties in ameliorating the impact of feedback may come through a return to the attitude formation literature. For example, in his comprehensive work on attitude formation and change, Cialdini (2001) reminds us that belief systems are based on multiple supporting cognitions. Initial commitment to a decision will often prompt the sprouting of new self-generated arguments consistent with that position. Thus, the eyewitness's identification decision itself may spur or reinforce a network of consistent self-statements and beliefs: "I have good memory for strangers," "The image in my mind is clear," "I would trust an eyewitness with the same experience." A confidence prophylactic likewise affects measures beyond the single judgment of confidence, and even inflates control group confidence (a "thought alone" effect) despite saying nothing about the accuracy of the identification (Wells & Bradfield, 1999). Confirming feedback thereby may prompt witnesses to construct a belief system in which the feedback is correct. Questions posed to the witness allow the expansion of cognitions to aspects perhaps not previously contemplated (e.g., view, attention, certainty, features of the face) and through causal chains not yet identified, as noted above.

Furthermore, the cue-accessibility framework relies on the notion that the witness is *actively* searching and evaluating information to arrive at reasonable judgments in response to the experimenter's questions. Yet, the network of beliefs that supports eyewitness memory is formed through a combination of deliberative thought and more automatic cognitive processes (such as priming) that foster an associative chain of compatible thoughts. Hence, a mitigation strategy that relies on active cognitive deliberation may be only partially successful in eliminating a belief structure that was formed through both deliberative and automatic cognitive processes. Indeed, the meta-analysis indicates a significant residual impact on witness responses (medium effect sizes) even when witnesses actively process information that should disarm the feedback. When such an intervention is at least partially successful, witnesses appear to have accepted the rationale for disregarding the feedback (e.g., "It was a mistake"). Indeed, a simple instruction that offers no information ("Ignore the feedback, rely on memory") completely fails to mitigate the feedback effect.

As per Cialdini's analysis of the low-ball technique (Cialdini, 2001, p. 89), attempts to undermine the central support of a belief system—such as an attack on credibility of the feedback—can be only somewhat effective, as the belief system is more firmly fixed with a scaffolding of beliefs that extends well beyond the feedback. This extensive belief system helps to maintain residual effects of the feedback even though a relatively successful mitigation strategy has been applied. Future research exploring how a witness's belief system grows in alignment with feedback—including broader witness-investigator communications ("You might just think on it a bit")—may be particularly relevant to the experience of real witnesses, who, compared with our laboratory participants, spend much more time in rumination about significant crime events and are more likely motivated toward self-persuasion.

The confirming versus disconfirming asymmetry. The meta-analysis shows a clear difference between confirming and disconfirming feedback in terms of the magnitude of their effects. The inflating effects of confirming feedback are much larger than the deflating effects of disconfirming feedback, especially for mistaken identifications from culprit-absent lineups. As currently construed, the cue accessibility conceptualization is silent on this consistent asymmentry between confirming and disconfirming feedback. However, a small

adjustment to the selective cue integration framework (Charman et al., 2010) might be able to explain the asymmetry post hoc. Recall that the selective cue integration framework has a "credibility check" in the hypothesized process. Specifically, when internal cues are weak the witness submits external sources to a credibility check before deciding whether to trust them. But why would confirming feedback be perceived by witnesses as more credible than disconfirming feedback? When witnesses make an identification from a lineup, they choose the person who they believe is most likely to be the culprit. Disconfirming feedback challenges that belief and perhaps this is just another example of people being more resistant to information that conflicts with their prior beliefs than to information that agrees with their prior beliefs (e.g., Anderson, Lepper, & Ross, 1980).

Another possibility, however, is that the particular way in which disconfirming feedback is manipulated in post-identification feedback experiments is responsible for the asymmetry. In the case of disconfirming feedback, the witness is told something that implies that a specific other member of the lineup was the culprit. Because witnesses tend to prefer the same lineup member in culprit-absent lineups (i.e., the one who looks most like the culprit), disconfirming feedback involves a potentially dubious claim, namely that the culprit was lineup member who looks less like the culprit than does the lineup member who was picked by the witness. If that explanation is correct, then the asymmetric effect sizes for confirming versus disconfirming feedback might be more of product of the specific manipulations of disconfirming feedback than it is a general phenomenon.

Future studies might explore the generality of the confirming/disconfirming asymmetry by using different operationalizations of the disconfirming feedback manipulation. For instance, would disconfirming feedback be weaker than confirming feedback if the disconfirming feedback told witnesses that none of the lineup members was the culprit? Another possibility is to create culprit-absent lineups in which two lineup members (e.g., Numbers 3 and 5) are *equally* similar to the culprit and if the witness picks number three they are told it was Number 5 (or vice versa). Our point is that the current literature cannot tell us whether the confirming/disconfirming asymmetry is a general phenomenon or whether it is specific to the particular way in which disconfirming feedback is operationalized in these experiments. If the confirming/disconfirming asymmetry holds up across different ways of manipulating disconfirming feedback, then theoretical conceptualizations of the post-identification feedback effect should try to incorporate the asymmetry into their accounts of the effect.

Retrospective certainty: The most important variable? The *Lawson* decision highlights a critical interest of eyewitness researchers: the appearance of reliability (confident testimony) without reliability itself (accuracy), or more directly put, the relation of confidence and accuracy (e.g., Sporer, Penrod, Read, & Cutler, 1995). The magnitude of the relation is highly variable and depends on a host of other variables such as the optimality of the viewing and testing conditions and the physical similarity between the culprit and the mistakenly identified individual. Much of the interest in the post-identification feedback effect is driven by the fact that certainty is readily and strongly influenced without a concomitant change in identification accuracy. Feedback distorts the relationship between eyewitness accuracy and confidence, with disconfirming feedback deflating the confidence of accurate witnesses as well as inflating confidence of inaccurate witnesses.

Whereas the wealth of eyewitness identification literature devoted to eyewitness identification certainty is understandable, it might also be a bit myopic. Witness self-reports of view and attention also influence perceived witness credibility (Bradfield & Wells, 2000). Accordingly, the fact that post-identification feedback maleates witnesses' reports of their view and attention is of considerable concern. And yet, outside of the post-identification feedback literature, almost every eyewitness identification study routinely measures witness certainty and almost no eyewitness identification study measures witnesses' reports of their views or attention. Two exceptions to this are studies by Bradfield et al. (2002) and by Smalarz and Wells (in press) that showed that, in the absence of feedback, witnesses' reports of their view and attention were significant indicators of whether they had made accurate or mistaken identifications.

Potential Limitations of the Literature

The post-identification feedback effect literature is very consistent and the effect sizes are large. Nevertheless, we note that there has been a lot of sharing of materials. Two videos, the "bomber on the roof" video and the "Target store security" video, constitute about half of all the witnessed events in published post-identification feedback experiments. There are an additional eight videos that have also been used and each shows a strong effect as well as one experiment using live exposure to a person. But, with only a couple of exceptions, the sampling of witnessed events across experiments is not particularly broad. Moreover, there has been no systematic manipulation of the characteristics of the witnessed event (e.g., exposure durations). This leaves open the question of whether there might be events (e.g., strongly encoded ones) that mitigate the effect.

Another potential limitation of the literature is that the feedback given to witnesses has almost exclusively been one of two types: lineup administrators telling witnesses that they identified the suspect or telling witnesses that a high percentage of other people had identified the same person that they had identified. Researchers should consider a broader set of manipulations. For example, what happens if the witness learns that the person he or she identified was arrested? Or, what happens if the witness is told of other evidence against the identified person (e.g., "You identified a guy who was found with the same amount of cash on him as was stolen")? Or, what if the witness is told about the absence of exculpatory evidence (e.g., "The guy you identified has no alibi")? Presumably, these are forms of feedback as well. And, according to current conceptualizations of the processes underlying the feedback effect, anything that tends to confirm witnesses' identification decisions should produce the post-identification feedback effect. But the research has not attempted to push these possible boundary conditions.

The post-identification feedback literature is grounded in a method that manipulates feedback as a single, one-time event. But in actual criminal cases, confirming feedback is often a series of

feedback events. This series might start with lineup administrator feedback but later include cowitness feedback (learning that another witness identified the same person), pretrial publicity feedback (e.g., media coverage indicating the person had been previously arrested for a similar offense), reinforcing feedback from pretrial interactions with a prosecutor ("Thank you for helping us solve this crime . . . now we have to get a conviction"), and so on. Does post-identification feedback cumulate? How far can this effect be pushed? For example, could the percent of mistaken witnesses with certainty of 80% or greater [see Figure 2] rise even higher to 60% or 90% if multiple instances of feedback were used? We also do not yet know how witnesses will respond to contradictory feedback. For example, how will a witness react if, after hearing confirming feedback, he or she then learns that the person identified had an iron-clad alibi?

Final Observations: A Disconnect Between the Memory System and the Legal System

Why do eyewitnesses not form better memories for these specific aspects of crime events and identification procedures? Wells and Quinlivan (2009a) posited that the "failure" of the cognitive system to lay a memory trace for the retrospective judgments required of eyewitnesses might simply be a functional and adaptive characteristic of a limited cognitive processing capacity organism. After all, the primary tasks at hand for the witness is to make sense of an unfolding crime event and to make a subsequent identification decision. If the cognitive system were to devote its limited resources to laying memory traces for these metacognitive judgments, performance on the primary task would be harmed. "From an evolutionary perspective, the individual who sees a bear and spends cognitive resources developing a memory for how good his view is of the bear or how much attention he is paying to the bear is more likely to be bear meat than he is to be one of our ancestors" (Wells & Quinlivan, 2009a, p. 1160). We probably should not be surprised that our cognitive system devotes little or no resources to laying memory records for these kinds of judgments that the legal system asks eyewitnesses to make. In fact, we are hard pressed to think of a situation outside of the specific needs of the legal system where people are required to give reliable retrospective reports on how much attention they paid, how good their view was, or how certain they were at the time of some decision.

The legal system further asks eyewitnesses to report memory based only on personal observation of the event at the time. Yet, recall is an ongoing constructive process that is guided by a person's state of knowledge at the time of retrieval (Loftus, 2005; Ross, 1990). Indeed, it is extremely difficult for a person to parse knowledge retrospectively based on *when* the information was acquired. Moreover, people tend to more quickly forget the source of information than the information itself (Brown, Deffenbacher, & Sturgill, 1977). And, people are largely wired to look for the best possible answer given all evidence available *at the time that a question is asked*. The report required of a witness seems to be a unique feature of the modern courtroom that arises for a specific purpose and it seems very unlikely that a cognitive module would have developed for making such judgments. Indeed, an important characteristic of these retrospective judgments asked by witnesses in the legal system is that they require the person to ignore what would normally be very relevant information (outcome feedback). In most everyday judgment tasks, people try to consider all relevant information in making retrospective judgments.

These inherent patterns of human cognitive processing greatly limit the likelihood that eyewitness performance will match legal expectations. Furthermore, we now know that witness retrospective memories for crime events and identification procedures are enormously influenced by even well-intentioned feedback from lineup administrators. The primary lesson of the post-identification feedback effect is that only way to know how certain the witness was at the time of the identification—to avoid the appearance of reliability without reliability itself—is to ask the witness about certainty at the time of the identification and prior to the contamination of post-identification influences. Our recommendation to address this problem is for double-blind lineup procedures that secure immediate witness reports of certainty and other testimony-relevant memory factors. We also recommend that identification procedures be videotaped (e.g., Kassin, 1998). Recent research emphasizes the critical importance of having the original confidence statement recorded so that triers of fact can adequately compare it with the (potentially inflated) confidence at trial (Douglass & Jones, 2013). Implementing these recommendations will increase the likelihood that eyewitness reports are probative (cf. *Oregon v. Lawson*, 2012) rather than reflections of a distorted memory construction process.

References

* Studies included in meta-analysis of feedback effects.

Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology, 39,* 1037–1049. doi:10.1037/h0077720

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

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

Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6, pp. 1–62). New York: Academic Press.

Bradfield, A. L., & Wells, G. L. (2000). The perceived validity of eyewitness identification testimony: A test of the five Biggers criteria. *Law and Human Behavior, 24,* 581–594. doi:10.1023/A:1005523129437

Bradfield, A., & Wells, G. L. (2005). Not the same old hindsight bias: Outcome information distorts a broad range of retrospective judgments. *Memory & Cognition, 33,* 120–130. doi:10.3758/BF03195302

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

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

Bushman, B. J., & Wells, G. L. (2001). Narrative impressions of the literature: The availability bias and the corrective properties of meta-analytic approaches. *Personality and Social Psychology Bulletin, 27,* 1123–1130. doi:10.1177/0146167201279005

*Charman, S. D., Carlucci, M., Vallano, J., & Gregory, A. (2010). The selective cue integration framework: A theory of post-identification

witness confidence assessment. *Journal of Experimental Psychology: Applied, 16,* 204–218. doi:10.1037/a0019495

*Charman, S. D., & Wells, G. L. (2008). Can eyewitnesses correct for external influences on their lineup identifications? The actual/counterfactual assessment paradigm. *Journal of Experimental Psychology: Applied, 14,* 5–20. doi:10.1037/1076-898X.14.1.5

*Charman, S. D., & Wells, G. L. (2012). The moderating effect of ecphoric experience on post-identification feedback: A critical test of the cues-based inference conceptualization. *Applied Cognitive Psychology, 26,* 243–250. doi:10.1002/acp.1815

Cialdini, R. B. (2001). *Influence: Science and practice* (4th ed.). Boston, MA: Allyn & Bacon.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Dixon, S., & Memon, A. (2005). The effect of post-identification feedback on the recall of crime and perpetrator details. *Applied Cognitive Psychology, 19,* 935–951. doi:10.1002/acp.1132

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

*Douglass, A. B., & McQuiston-Surrett, D. M. (2006). Post-identification feedback: Exploring the effects of sequential photospreads and eyewitnesses' awareness of the identification task. *Applied Cognitive Psychology, 20,* 991–1007. doi:10.1002/acp.1253

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

Douglass, A. B., & Steblay, N. (2006). Memory distortion in eyewitnesses: A meta-analysis of the post-identification feedback effect. *Applied Cognitive Psychology, 20,* 859–869. doi:10.1002/acp.1237

*Douglass, A. B., Brewer, N., & Semmler, C. (2010). Moderators of post-identification feedback effects on eyewitnesses' memory reports. *Legal and Criminological Psychology, 15,* 279–292. doi:10.1348/135532509X446337

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

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 288–299. doi:10.1037/0096-1523.1.3.288

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

Garrioch, L., & Brimacombe, C. A. E. (2001). Lineup administrators' expectations: Their impact on eyewitness confidence. *Law & Human Behavior, 25,* 299–314. doi:10.1023/A:1010750028643

Gerrard, M., Gibbons, F. X., & Houlihan, A. E., Stock, M. L., & Pomery, E. A. (2008). A dual-process approach to health risk decision making: The prototype willingness model. *Developmental Review, 28,* 29–61. doi:10.1016/j.dr.2007.10.001

*Hafstad, G. S., Memon, A., & Logie, R. (2004). Post-identification feedback, confidence and recollections of witnessing conditions in child witnesses. *Applied Cognitive Psychology, 18,* 901–912. doi:10.1002/acp.1037

Horry, R., Halford, P., Brewer, N, Milne, R., & Bull, R. (in press). Archival analyses of eyewitness identification test outcomes: What can they tell us about eyewitness memory? *Law and Human Behavior*.

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

Ionnadis, J. P. A. (2008). Why most discovered true associations are inflated. *Epidemiology, 19,* 640–648.

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

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

*Lampinen, J., Scott, J., Pratt, D., Leding, J. K., & Arnal, J. D. (2007). "Good, you identified the suspect. . .but please ignore this feedback:" Can warnings eliminate the effects of post-identification feedback? *Applied Cognitive Psychology, 21,* 1037–1056. doi:10.1002/acp.1313

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory, 12,* 361–366. doi:10.1101/lm.94705

Maclean, C. L., Brimacombe, C. A. E., Allison, M., Dahl, L. C., & Kadlec, H. (2011). Post-identification feedback effects: Investigators and evaluators. *Applied Cognitive Psychology, 25,* 739–752. doi:10.1002/acp.1745

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

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

Neil v. Biggers, 409 US 188 (1972).

*Neuschatz, J. S., Lawson, D. S., Fairless, A. H., Powers, R. A., Neuschatz, J. S., Goodsell, C. A., & Toglia, M. P. (2007). The mitigating effects of suspicion on post-identification feedback and on retrospective eyewitness memory. *Law and Human Behavior, 31,* 231–247. doi:10.1007/s10979-006-9047-7

*Neuschatz, J. S., Preston, E. L., Burkett, A. D., Toglia, M. P., Lampinen, J. M., Neuschatz, J. S., . . . Goodsell, C. A. (2005). The effects of post-identification feedback and age on retrospective eyewitness memory. *Applied Cognitive Psychology, 19,* 435–453. doi:10.1002/acp.1084

New Jersey v. Henderson, 2011, Report of the Special Master.

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

Oregon v. Lawson, SC S059306, 2012.

*Quinlivan, D. S., Neuschatz, J. S., Douglass, A. B., Wells, G. L., & Wetmore, S. A. (2012). The effect of post-identification feedback, delay, and suspicion on accurate eyewitnesses. *Law and Human Behavior, 36,* 206–214. doi:10.1037/h0093970

*Quinlivan, D. S., Neuschatz, J. S., Jimenez, A., Cling, A. D., Douglass, A., & Goodsell, C. A. (2009). Do prophylactics prevent inflation? Post-identification feedback and the effectiveness of procedures to protect against confidence-inflation in ear-witnesses. *Law and Human Behavior, 33,* 111–121. doi:10.1007/s10979-008-9132-1

*Quinlivan, D. S., Wells, G. L., & Neuschatz, J. S. (2010). Is manipulative intent necessary to mitigate the eyewitness post-identification feedback effect? *Law and Human Behavior, 34,* 186–197. doi:10.1007/s10979-009-9179-7

Rodriguez, D. N., & Berry, M. A. (2010). System and estimator variables, eyewitness confidence, and the post-identification feedback effect. *American Journal of Forensic Psychology, 28,* 17–37.

Rosenthal, R. (1964). The effect of the experimenter on the results of psychological research. *Bulletin of the Maritime Psychological Association, 13*(1), 1–39.

Rosenthal, R. (1991). Teacher expectancy effects: A brief update 25 years after the Pygmalion experiment. *Journal of Research in Education, 1,* 3–12.

Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences, 1,* 377–415. doi:10.1017/S0140525X00075506

Ross, L. (1990). Recognizing the role of construal processes. In I. Rock (Ed.), *The legacy of Solomon Asch: Essays in cognition and social psychology* (pp. 77–96). Hillsdale, NJ: Erlbaum, Inc.

Schooler, J. (2011). Unpublished results hide the decline effect. *Nature, 470,* 437. doi:10.1038/470437a

*Semmler, C., Brewer, N., & Wells, G. L. (2004). Effects of post-identification feedback on eyewitness identification and nonidentification. *Journal of Applied Psychology, 89,* 334–346. doi:10.1037/0021-9010.89.2.334

*Skagerberg, E. M., & Wright, D. B. (2009). Susceptibility to post-identification feedback is affected by source credibility. *Applied Cognitive Psychology, 23,* 506–523. doi:10.1002/acp.1470

*Smalarz, L., & Wells, G. L. (in press). Post-identification feedback to eyewitnesses impairs evaluators' abilities to discriminate between accurate and mistaken testimony. *Law and Human Behavior*.

Sporer, S., Penrod, S., Read, D., & Cutler, B. L. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin, 118,* 315–327. doi:10.1037/0033-2909.118.3.315

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

Wells, G. L. (1988). *Eyewitness identification: A system handbook*. Toronto, Canada: Carswell Legal Publications.

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

Wells, G. L. (1995). Scientific study of witness memory: Implications for public and legal policy. *Psychology, Public Policy, and Law, 1,* 726–731.

*Wells, G. L., & Bradfield, A. L. (1998). "Good, you identified the suspect:" Feedback to eyewitnesses distorts their reports of the witnessing experience. *Journal of Applied Psychology, 83,* 360–376. doi:10.1037/0021-9010.83.3.360

*Wells, G. L., & Bradfield, A. L. (1999). Distortions in eyewitnesses' recollections: Can the post-identification feedback effect be moderated? *Psychological Science, 10,* 138–144. doi:10.1111/1467-9280.00121

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

*Wells, G. L., Olson, E., & Charman, S. (2003). Distorted retrospective eyewitness reports as functions of feedback and delay. *Journal of Experimental Psychology: Applied, 9,* 42–52. doi:10.1037/1076-898X.9.1.42

Wells, G. L., & Quinlivan, D. S. (2009a). The eyewitness post-identification feedback effect: What is the function of flexible confidence estimates for autobiographical events? *Applied Cognitive Psychology, 23,* 751–762. doi:10.1002/acp.1616

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

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

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

Wells, G. L., Steblay, N. K., & Dysart, J. E. (2013). *Double-blind photolineups using actual eyewitnesses: An experimental test of a sequential versus simultaneous lineup procedure*. Manuscript under review.

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

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

Received August 29, 2013
Revision received November 8, 2013
Accepted November 14, 2013 ■