Law and Human Behavior
2014, Vol. 38, No. 2, 194–202


0147-7307/14/$12.00 DOI: 10.1037/lhb0000067

# Post-Identification Feedback to Eyewitnesses Impairs Evaluators' Abilities to Discriminate Between Accurate and Mistaken Testimony

Laura Smalarz and Gary L. Wells
Iowa State University

Giving confirming feedback to mistaken eyewitnesses has robust distorting effects on their retrospective judgments (e.g., how certain they were, their view, etc.). Does feedback harm evaluators' abilities to discriminate between accurate and mistaken identification testimony? Participant-witnesses to a simulated crime made accurate or mistaken identifications from a lineup and then received confirming feedback or no feedback. Each then gave videotaped testimony about their identification, and a new sample of participant-evaluators judged the accuracy and credibility of the testimonies. Among witnesses who were not given feedback, evaluators were significantly more likely to believe the testimony of accurate eyewitnesses than they were to believe the testimony of mistaken eyewitnesses, indicating significant discrimination. Among witnesses who were given confirming feedback, however, evaluators believed accurate and mistaken witnesses at nearly identical rates, indicating no ability to discriminate. Moreover, there was no evidence of overbelief in the absence of feedback whereas there was significant overbelief in the confirming feedback conditions. Results demonstrate that a simple comment following a witness' identification decision ("Good job, you got the suspect") can undermine fact-finders' abilities to discern whether the witness made an accurate or a mistaken identification.

*Keywords:* eyewitness identification, post-identification feedback, eyewitness testimony, overbelief



Eyewitnesses who take the stand to testify about the identity of a perpetrator are one of the most valuable sources of evidence in the criminal justice system. Often, eyewitness accounts are the only evidence available to help identify and prosecute suspects, such as in cases of drive-by shooting, robbery, assault, and others. Consequently, it has been of longstanding concern to legal scholars and psychological scientists whether triers-of-fact (judges and jurors) are capable of appropriately evaluating the accuracy of eyewitness accounts.

The advent of forensic DNA-testing in the 1990s injected a new urgency into the question regarding fact-finders' abilities to discriminate between accurate and mistaken eyewitness-identification testimony, because 75% of the individuals who were wrongfully convicted and later exonerated with DNA testing were cases of mistaken eyewitness identification (innocenceproject.org, 2013). Much of the empirical work tended to focus on how little triers-of-fact seem to know about the workings of memory to explain why mistaken eyewitness identification testimony is often believed (see meta-analysis by Desmarais & Read, 2011). In recent years, however, a complementary argument has emerged that places less blame on triers-of-fact and instead focuses more on the idea that the legal system might be artificially enhancing the credibility of mistaken eyewitnesses through postidentification feedback confirming witnesses' identification decisions. Numerous studies have now shown that false eyewitness confidence (being highly confident or certain and yet mistaken) is easily created by giving witnesses confirming feedback following a mistaken identification (e.g., "Good, you identified the suspect"). This postidentification feedback effect, first shown by Wells and Bradfield (1998), also occurs among children (e.g., Hafstad, Memon, & Logie, 2004), the elderly (e.g., Neuschatz et al., 2005), and for lineup rejections as well as identifications (e.g., Semmler, Brewer & Wells, 2004). Moreover, the postidentification feedback effect is both a large effect (see meta-analysis by Douglass & Steblay, 2006) and occurs with actual eyewitnesses to serious crimes (Wright & Skagerberg, 2007).

Giving confirming feedback to mistaken eyewitnesses not only gives witnesses false confidence but it also causes them to recall that their view was better than it was and recall that they paid more attention to the culprit, and it inflates several other self-report judgments by the witnesses. Importantly, the effects of feedback are observed on witnesses' *retrospective* judgments (e.g., "How certain were you *at the time* of your identification?" How good was the view you had of the culprit?), which concern their recollections of matters that *preceded* the feedback and thereby indicate distortion. The fact that confirming postidentification feedback distorts these retrospective judgments is potentially very worrisome. The empirical evidence indicates that people evaluate eyewitnesses' identification testimonies by relying on cues such how confident witnesses are, what witnesses say about their viewing conditions, and how much attention witnesses say they paid during the witnessing episode (Bradfield & Wells, 2000). Moreover, there is evidence that giving confirming feedback to mistaken eyewitnesses leads witnesses to give more persuasive testimony that,

This article was published Online First December 16, 2013.

Laura Smalarz and Gary L. Wells, Psychology Department, Iowa State University.

We thank Amy Douglass for sharing some materials that were adapted for use in the current research. This research was supported by an AP-LS Student Grant-in-Aid awarded to Laura Smalarz and a National Science Foundation Grant awarded to Gary Wells.

Correspondence concerning this article should be addressed to Laura Smalarz, Psychology Department, West 112 Lagomarcino Hall, Iowa State University, Ames, IA 50011. E-mail: lsmalarz@iastate.edu

when shown to participant-jurors, leads participant-jurors to higher rates of believing these mistaken-identification witnesses (Douglass et al., 2010; MacLean et al., 2011).

Our interest is similar to the Douglass et al. (2010) and MacLean et al. (2011) studies in one respect but importantly distinct in another respect. It is similar in the sense that we wanted to know how postidentification feedback affects the perceived credibility of eyewitness-identification testimony when the witnesses testify in their own words. But the current work is distinct in the sense that the primary question is what postidentification feedback does to the ability of testimony evaluators to *discriminate* between accurate and mistaken eyewitness-identification testimony. Both the Douglass et al. and the MacLean et al. studies used only mistaken eyewitnesses. Discrimination refers to the ability of testimony evaluators to believe accurate witnesses more than they believe mistaken eyewitnesses, which can only be tested by using both mistaken and accurate eyewitnesses. If confirming feedback is given to both mistaken eyewitnesses and to accurate eyewitnesses, it is possible for confirming feedback to increase the rate at which evaluators believe mistaken eyewitnesses without affecting discrimination. Specifically, discrimination would be unaffected if confirming feedback inflated the perceived credibility of accurate eyewitnesses as much as it inflates the perceived credibility of mistaken eyewitnesses.

Although no prior study has tested the question of whether confirming feedback to eyewitnesses harms testimony evaluators' abilities to discriminate between accurate and mistaken eyewitness-identification testimony, there are some reasons to believe that it would. Specifically, the cue-accessibility conceptualization states that witnesses will use external cues (such as feedback) to retrospectively infer testimony-relevant answers (e.g., about their certainty, view, attention) to the extent that internal cues are weak (Wells & Bradfield, 1999). A sense of actual recognition (high ecphoric similarity) has been posited as an internal cue available to accurate witnesses, thereby making accurate witnesses less influenced by external cues such as feedback (Bradfield et al., 2002; Charman & Wells, 2012). Consistent with this conceptualization, Bradfield et al. found that giving confirming feedback to both accurate and mistaken eyewitnesses reduced the certainty-accuracy relation because confirming feedback inflated the confidence of the mistaken eyewitnesses more than it inflated the confidence of the accurate eyewitnesses. Although Charman and Wells did not find a statistically significant reduction in the confidence-accuracy relation following confirming feedback, they did find that disconfirming feedback affected mistaken eyewitnesses more than it affected accurate witnesses, which provided at least provisional support for the idea that identification accuracy moderates feedback effects. In summary, there are theoretical reasons to believe that accuracy will moderate the effects of feedback on testimony evaluations and, although evidence supporting the proposition is mixed, we predicted that confirming feedback will increase the perceived credibility of mistaken eyewitnesses more than it increases the perceived credibility of accurate eyewitnesses.

## Testimony Versus Likert-Type Scale Measures

Adding some uncertainty to our prediction is the fact that postidentification feedback studies to date have almost exclusively measured feedback effects using pencil-and-paper or computer measures that constrain the witnesses to report their responses on Likert-type scales. Two exceptions, already mentioned earlier, had witnesses give testimony-like statements that were videotaped and shown to evaluators who made judgments about the believability of the testimony (Douglass et al., 2010; MacLean et al., 2011). However, those studies used only mistaken eyewitnesses and therefore could not address the question of whether confirming feedback harms evaluators' abilities to discriminate between accurate and mistaken eyewitness testimony.

How well do the postidentification feedback questions used in the extant literature capture the dynamics of oral testimony? Are those Likert-type measures good and complete proxies for evaluators' judgments about the credibility of the witness? The questions used in the postidentification feedback literature cover a relatively broad range, including certainty, view, attention, willingness to testify, ability to make out details of the face, and so on. But, oral eyewitness-testimony is characterized by communication dynamics that might not be captured by responses on Likert-type scales, such as narrative descriptions of the event, facial expressions, pauses, enthusiasm, verbal qualifiers, shoulder shrugs, fluency, tone of voice, and so on. Moreover, testimony is a more public, social display of behavior than are pencil-and-paper scale responses and this public accountability has the potential to be qualitatively different from what is captured by Likert-type scale responses. For example, manipulations that increase accountability have been shown to induce more complex thought about decisions that are yet to be made (Tetlock, 1983) and tend to induce bolstering and self-justification processes about decisions that have already been made (Kiesler, 1971; Tetlock, Skitka & Boettger, 1989). And in fact, research has demonstrated that bolstering occurs among witnesses who expect their testimony to be challenged during cross-examination (Wells, Ferguson, & Lindsay, 1981). The possibility that the Likert-type scale questions typically asked in postidentification feedback studies fail to capture some of the dynamics of oral testimony raises the possibility that the moderating role of witness accuracy on the feedback effect might not be exactly the same for Likert-type responses as it is for evaluations of oral testimony. Of course, in the end, it is testimony that courts use to make their evaluations of witness credibility, not Likert-type scale responses. After a talk given by the second author of the current article, one prosecutor remarked, "I don't care what a witness circles on a piece of paper . . . I only care how they testify."

## Overview

We tested the idea that giving confirming feedback to accurate and mistaken eyewitnesses would harm the abilities of testimony evaluators to discriminate between accurate and mistaken testimony. The extended version of the cue-accessibility conceptualization, which postulates that eyewitness accuracy moderates the effect of postidentification feedback (Bradfield et al., 2002; Charman & Wells, 2012), was the foundation for our expectation that confirming feedback would serve to diminish differences in testimony credibility between accurate and mistaken eyewitnesses. Moreover, although we expected witnesses' Likert-type scale responses to the standard postidentification feedback questions to be similar in many respects to evaluators' ratings of witnesses' oral testimonies, we considered that some differences might emerge.

Our general method was similar to the two-phase eyewitness paradigm (first used by Wells, Lindsay, & Ferguson, 1979) in which witnesses make identification decisions, give testimony about their identifications, and evaluators later view the testimonies and make judgments about the accuracy of the witnesses' identifications.

## Method

### Overview

We used a two-phase paradigm in which witnesses made either accurate or mistaken identifications and then were randomly assigned to receive confirming feedback or no feedback. After answering the traditional postidentification feedback measures, all witnesses were videotaped giving testimony about their identifications. In Phase 2, participant-evaluators observed the videos and made judgments about the accuracy/inaccuracy of the witnesses along with other judgments about the testimony.

### Phase 1: Witness Identifications and Testimony

Participants ($n = 140$) were university students who received course credit for their participation. Upon entering the lab, participants were told that the study was designed to investigate people's tendencies to form impressions of others after having viewed them for a short time. Participants viewed a one minute and 28 second video of a simulated crime in which a man at an airport switched his bag with another passenger's bag. The video was clear, in color, and provided multiple views of the culprit's face. After viewing the video, participant-witnesses (hereafter called witnesses) were told that the bag left behind by the man contained a bomb and that the purpose of the study was actually to see whether they could identify the culprit from a photo lineup.

Like other postidentification feedback studies, accuracy was manipulated by including or removing the culprit from the lineup and using a procedure that induces all witnesses to make an identification. The procedure implies that the culprit is present ("Try to identify the person who switched the bags") and excludes an explicit "Not there" option. As a result, all witnesses make an identification, most witnesses who receive a culprit-present lineup make an accurate identification, and all witnesses who receive a culprit-absent lineup make a mistaken identification. Studies have shown that the postidentification feedback effect for identifying witnesses occurs even if the witnesses are given unbiased instructions and allowed to make no identification (Semmler et al., 2004; Wright & Skagerberg, 2007). However, for efficiency reasons, getting all witnesses to make an identification has been a standard procedure in postidentification feedback studies because witnesses who make no identification would simply have to be discarded. We consider possible implications and limitations of this method of manipulating accuracy in the Discussion section.

The culprit-present lineup consisted of a photo of the culprit plus five photos of individuals who matched the general description of the culprit. We intentionally used a video and culprit-present lineup that pilot data shows produce a high rate of accurate identification. This is consistent with our interest in efficiency, as identifications of fillers from the culprit-present lineup are wasted. The culprit-absent lineup consisted of the same five filler photos that were used in the culprit-present lineup; the culprit's photo was simply removed. As expected, all witnesses made an identification. Eighty-four percent of witnesses who viewed the culprit-present lineup accurately identified the culprit (accurate witnesses; $n = 64$), and 100% of witnesses who viewed the culprit-absent lineup made mistaken identifications (mistaken witnesses; $n = 64$). After their identifications, witnesses were randomly assigned to receive confirming feedback ("Good job! You got the guy.") or no feedback. The feedback was delivered orally by the experimenter who

Table 1
*Witnesses' Self-Report Questions*

| Measure | Question | Scale |
|---|---|---|
| Certainty | At the time you identified the person from the photo lineup, how certain were you that the person you identified from the photo lineup was the person you saw in the video? | 10% (*not at all certain*) to 100% (*totally certain*), in 10% intervals |
| View | How good a view did you get of the person in the video? | 0 (*very poor*) to 10 (*very good*) |
| Face | How well were you able to make out specific features of the person's face from the video? | 0 (*not at all*) to 10 (*very well*) |
| Attention | How much attention were you paying to the face of the person in the video while viewing the tape? | 0 (*none*) to 10 (*my total attention*) |
| Basis | To what extent do you feel that you had a good basis to make an identification? | 0 (*no basis at all*) to 10 (*a very good basis*) |
| Ease | How easy or difficult was it for you to figure out which person in the photo lineup was the person from the video? | 0 (*extremely difficult*) to 10 (*extremely easy*) |
| Time | From the time the lineup started, how much time do you estimate it took you to make an identification? | 0 (*I needed almost no time*) to 10 (*I had to look at the photos for a long time*) |
| Willing | On the basis of your memory of the person in the video, how willing would you have been to testify in court that the person you identified was the same person you saw in the video? | 0 (*not at all willing*) to 10 (*totally willing*) |
| Strangers | Generally, how good is your recognition memory for faces of strangers you have encountered on only one prior occasion? | 0 (*very poor*) to 10 (*excellent*) |
| Image | How clear is the image you have in your memory of the person you saw in the video? | 0 (*not at all clear*) to 10 (*very clear*) |

was serving as the lineup administrator. Witnesses then answered the traditional self-report measures of the postidentification feedback effect (see Table 1).

Witnesses were subsequently directed to another room where they were met by a new experimenter who was blind to their experimental condition. The experimenter gave the witnesses an agreement form in which witnesses were asked to consent to being videotaped during a testimony interview. All witnesses consented to being videotaped. The experimenter then began the testimony interview, working from a scripted series of questions that always allowed witnesses to use their own words and never cut off the witnesses' answers. In the initial questioning, witnesses were asked to describe in as much detail as possible what they witnessed in the video. Witnesses typically provided narrative descriptions including details about the airport setting, bystanders at the scene, actions taken by the culprit before, during, and after switching the bags, and the physical appearance of the culprit. After these narrative event descriptions, witnesses gave testimony responses to a series of questions that were similar to standard postidentification feedback questions (e.g., How much attention were you paying to the face of the person who switched the bags?). A full text of the testimony-interview questions and the video and lineup materials are available from the authors upon request.

### Phase 2: Evaluator Judgments of Testimony

Each participant-evaluator ($n = 64$) viewed the videotaped testimony of four eyewitnesses (one from each condition, in random order), thereby yielding two independent evaluations of the testimony for each of the 128 videos. The evaluators were not informed that some of the witnesses had received confirming feedback. To deter strategic guessing regarding the accuracy of the witnesses, evaluators were led to believe that they would be viewing the testimonies of five witnesses and that the actual accuracy of the witnesses was unknown, therefore making it possible that they would see all inaccurate witnesses, all accurate witnesses, or some combination of accurate and inaccurate witnesses. After viewing each video, evaluators indicated on the computer whether they believed the witness had made an accurate or inaccurate identification and they answered additional questions about their perceptions of the witness' testimony (see list in Table 2).

## Results

### Phase 1: Witnesses' Self-Reports

We first analyzed the traditional witness self-report measures (questions are in Table 1) using a multivariate analysis of variance (MANOVA), which constituted a single measure of witnesses' self-reported reliability. Results for the univariate analyses for each individual measure are available from the first author upon request. In line with past research, accurate witnesses had a higher standing on self-reported reliability than did mistaken witnesses, $F(10, 115) = 5.86$, $p < .001$ partial eta-squared $= .34$, and feedback significantly inflated witnesses' self-reported reliability, $F(10, 115) = 2.74$, $p = .005$ partial eta-squared $= .19$ (see Figure 1). We did not observe a significant accuracy-by-feedback interaction on witnesses' self-reported reliability, $F(10, 115) = 1.13$, $p = .35$ partial eta-squared $= .09$. We return to this result in an analysis of the ability of the traditional feedback measures to account for the pattern of evaluators' belief judgments.

### Phase 2: Evaluators' Judgments of Witnesses' Testimonies

Our primary dependent measure was whether or not evaluators believed that the eyewitness had made an accurate identi-

Table 2
*Evaluators' Testimony Judgment Questions*

| Measure | Question | Scale |
|---|---|---|
| Belief | Do you think that the witness' identification from the photo lineup of the man who switched the bags was an accurate identification or an inaccurate identification? | Accurate -or- Inaccurate |
| Convincing | How convincing was the witness? | 1 (*Not at all*) to 10 (*Totally*) |
| Confidence felt | How confident do you believe the witness felt that the person s/he identified was the person who switched the bags? | 1 (*Not at all*) to 10 (*Totally*) |
| Confidence portrayed | How confident do you think the witness portrayed him or herself to be? | 1 (*Not at all*) to 10 (*Totally*) |
| View of culprit | How good of a view do you think the witness got of the man who switched the bags? | 1 (*Very poor*) to 10 (*Very good*) |
| Mental image of culprit | How clear of a mental image do you think the witness had in memory of the person who switched the bags? | 1 (*Not at all*) to 10 (*Very*) |
| Event description narrative | How accurate of a description do you think the witness provided about what s/he witnessed in the video? | 1 (*Not at all*) to 10 (*Totally*) |
| Attention paid | How much attention do you think the witness was paying when s/he witnessed the event? | 1 (*No attention*) to 10 (*Total attention*) |
| Ability to recognize strangers | How good do you think the witness is at remembering faces of strangers? | 1 (*Very poor*) to 10 (*Excellent*) |
| Additional evidence (reverse coded as evidence sufficiency) | If you were a juror at trial, how much additional evidence would you need to convict the person who was identified by the witness as the man who switched the bags? | 1 (*No additional evidence*) to 10 (*A lot of additional evidence*) |
| Confidence reported | During the witness' testimony, how confident did the witness report having been in his/her identification? | 1 (*Not at all*) to 10 (*Totally*) |

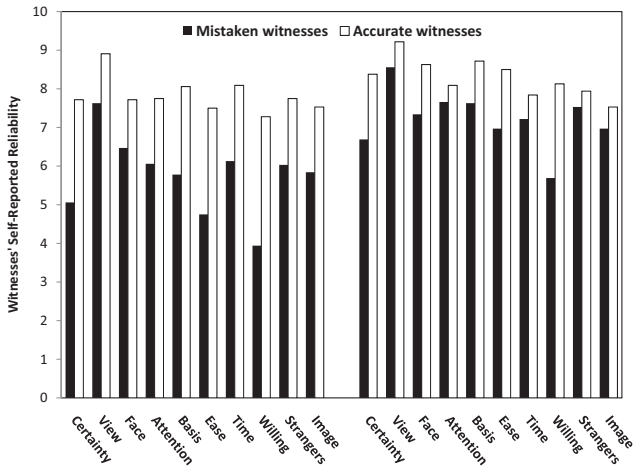


*Figure 1.* Witnesses' self-reported reliability. Time measure is reverse-coded. See Table 1 for questions.

fication. Because belief judgments were nested within evaluators (each evaluator evaluated four witnesses), observations were not independent. Therefore, a multilevel analysis was conducted using SAS PROC GLIMMIX for binary outcomes in which evaluators were identified as random effects to account for the nonindependence of the dependent variable. Because each evaluator evaluated four eyewitnesses, we believed it was important to first assess whether evaluators' abilities to discern whether a witnesses made an accurate or mistaken identification (i.e., the accuracy of evaluators' judgments) improved as they viewed more witnesses. Accordingly, we constructed a model in which we assessed the effects of witness order on the accuracy of evaluators' judgments. The effect for witness order was nonsignificant, $F(3, 186) = 1.05$, $p = .37$. Therefore, we omitted this factor from the remaining analyses and used a model that included the main effects of accuracy and feedback and the accuracy-by-feedback interaction as predictors of evaluators' belief judgments. Effect size estimates for numeric scale measures are reported using Cohen's *d*, whereas effect size estimates for differences in proportions are reported using Cohen's *h*. A *d* or an *h* of 0.2, 0.5, and 0.8 correspond to a small, medium, and large effect size, respectively (Cohen, 1988).

Our main research question was whether feedback interferes with evaluators' abilities to discriminate between accurate and mistaken identification testimony. Therefore, we first tested for an interaction between accuracy and feedback on evaluators' belief that the witness made an accurate identification. The interaction was significant, $F(1, 189) = 6.81$, $p = .01$. We conducted two simple-effects tests to identify the source of the interaction. As shown in Figure 2, when witnesses were not given feedback, evaluators were significantly more likely to believe accurate witnesses than they were to believe mistaken witnesses (70.3% vs. 35.9%, respectively), $F(1, 189) = 14.53$, $p < .001$, odds ratio (OR) = 4.22, 95% CI (OR) = [2.00, 8.90], $h = .70$. In contrast, when witnesses had received feedback, belief rates of accurate versus mistaken witnesses were nearly identical (64.1% vs. 62.5%, respectively), $F(1, 189) = .03$, $p = .85$, OR = 1.07, 95% CI (OR) = [.52, 2.21], $h = .02$. Thus, feedback eliminated evaluators' abilities to discriminate between the testimony of accurate and mistaken eyewitnesses.

We were also interested in examining the rates of false positives versus false negatives in each of the feedback conditions. In the absence of feedback, there was no evidence of overbelief as indicated by the fact that false positives (believing a mistaken witness) and false negatives (disbelieving an accurate witness) were not significantly different (36% and 30%, respectively), $F(1, 189) = .57$, $p = .45$, OR = 1.33, 95% CI (OR) = [.63, 2.80], $h = .13$. In contrast, false positives (63%) were considerably higher than false negatives (36%) in the confirming feedback conditions, indicating overbelief, $F(1, 189) = 8.81$, $p = .003$, OR = 2.97, 95% CI (OR) = [1.44, 6.13], $h = .55$.

We next analyzed evaluators' other judgments of the eyewitnesses' testimonies using a mixed-effects model on a composite measure of all of the testimony-relevant judgments (see list of questions in Table 2; means in Figure 3). The composite measure, which constituted a general measure of testimony credibility, was created by averaging across all of the testimony judgments, with the *additional evidence* measure being reverse-coded to create a measure of "evidence sufficiency." Hence, higher values on the composite measure correspond to higher assessments of testimony credibility. Evaluators were again identified as random effects and accuracy, feedback, and the accuracy-by-feedback interaction were identified as fixed effects. Replicating the pattern of belief rates, evaluators perceived the testimony of accurate witnesses as being significantly more credible than the testimony of mistaken witnesses, $F(1, 112) = 31.15$, $p < .001$, $d = .62$, 95% CI ($d$) = [.37, .87], and feedback significantly inflated evaluators' credibility judgments, $F(1, 94) = 18.12$, $p < .001$, $d = .33$, 95% CI ($d$) = [.08, .58]. There was also a significant accuracy-by-feedback interaction on evaluators' assessments of the witnesses' testimonies, $F(1, 188) = 9.52$, $p = .002$. Specifically, in the no-feedback conditions, evaluators judged the testimonies of accurate witnesses to be more credible than the testimonies of mistaken eyewitnesses, $t(63) = 5.35$, $p < .001$, $d = .67$, 95% CI ($d$) = [.42, .92]. When the witnesses had received feedback, however, evaluators' credibility judgments of accurate and mistaken eye-

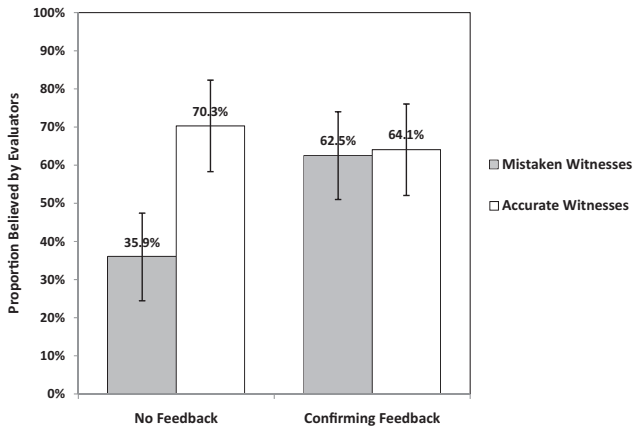


*Figure 2.* Belief of witnesses as functions of accuracy and feedback with 95% confidence intervals.

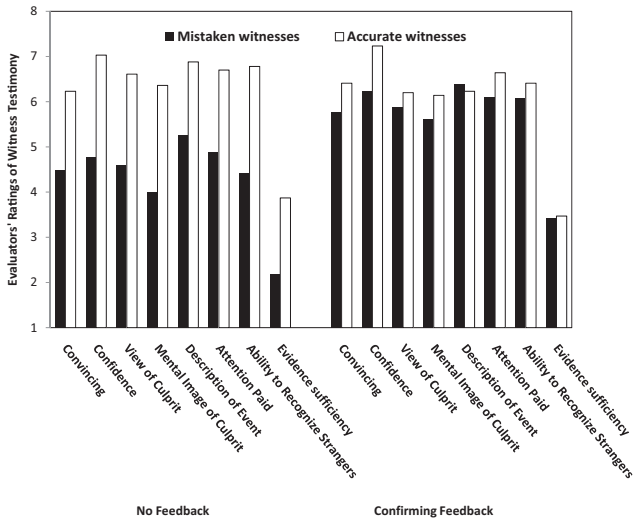


*Figure 3.* Evaluators' testimony judgments. Confidence measure reflects the average of evaluators' judgments of Confidence Felt, Confidence Portrayed, and Confidence Reported, which all behaved similarly during individual analyses (See Table 2 for questions).

witnesses did not differ, $t(63) = 1.58$, $p = .12$, $d = .20$, 95% CI ($d$) = [−.05, .45].

## Can the Traditional Feedback Measures Account for the Pattern of Evaluators' Belief Judgments?

We were also interested in knowing whether the traditional witness self-report measures could account for the pattern of data on evaluators' belief judgments. This addresses the fundamental question of whether the credibility of witnesses' testimonies is a simple reflection of how they answer the standard postidentification feedback questions. The fact that the accuracy-by-feedback interaction was significant for testimony belief but not for the standard set of postidentification feedback questions suggests that the answer might be "no." Therefore, we created a model in which we included the standard set of witnesses' self-report measures along with the main effects of accuracy, feedback, and the accuracy-by-feedback interaction as predictors of evaluators' belief in testimony. Controlling for witnesses' self-report measures eliminated the main effects of accuracy and feedback on testimony belief rates, $F$s (1, 179) ≤ .49, $p$s ≥ .48, but did not eliminate the accuracy-by-feedback interaction, $F(1, 179) = 4.12$, $p = .04$. This finding, along with the observed interaction in evaluators' belief judgments but not in witnesses' self-reports, indicates that the traditional self-report measures did not fully account for the pattern of the evaluators' belief judgments.

Eyewitnesses' self-report measures might have failed to account for the interaction effect in evaluators' belief judgments because the traditional measures do not necessarily capture the richness of eyewitnesses' testimony narratives. Specifically, at the beginning of the testimony interviews, witnesses provided free-recall narrative descriptions of what they witnessed in the crime video. The unstructured nature of the narratives allowed for considerable variation in verbal and nonverbal elements such as level of exuberance, insertion of details, and apparent willingness to share account recollections. These testimony dynamics—while readily perceptible to evaluators—are not tapped by the traditional self-report measures. Hence, it is possible that the effects of feedback were transmitted to evaluators through these verbal and nonverbal communication dynamics, which have the potential to operate independently from witnesses' answers to the questions themselves.

To test this possibility, we assessed evaluators' impressions of the witnesses' narrative descriptions by asking evaluators to rate the quality of the eyewitnesses' descriptive accounts ("How accurate of a description do you think the witness provided about what s/he witnessed in the video?"). Evaluators' ratings of witnesses' narrative event descriptions followed a pattern similar to the pattern observed in evaluators' belief judgments, such that they distinguished reliably between accurate and mistaken eyewitness accounts in the absence of feedback, $t(63) = 3.67$, $p < .001$, $d = .46$, 95% CI ($d$) = [.21, .71] but not when feedback had been delivered, $t(63) = .32$, $p = .75$, $d = -.04$, 95% CI ($d$) = [−.21, .29]. Therefore, we created a new model in which we included evaluators' narrative description ratings along with all of the variables that were included previously (the main effects of accuracy and feedback, the accuracy-by-feedback interaction, and all of the witnesses' self-report measures) as predictors of evaluators' belief judgments. Not only did controlling for evaluators' ratings of the witnesses' event descriptions eliminate the accuracy-by-feedback interaction, $F(1, 178) = 1.45$, $p = .23$, but the narrative-description ratings variable was the only variable that significantly predicted evaluators' belief judgments, $F(1, 178) = 29.91$, $p < .001$; all other variables $F$s(1,178) ≤ 2.58, $p$s ≥ .11. This suggests an important role of the narrative event descriptions in determining evaluators' propensities to believe the testimony of an eyewitness. Because narrative event descriptions are not part of the traditional postidentification feedback questions, this finding lends support to the idea that the process by which feedback influences eyewitness testimony is richer and more complex than can be detected by the traditional measures of the postidentification feedback effect.

## Discussion

Testimony evaluators were able to significantly discriminate between accurate and mistaken witnesses when the witnesses had not received confirming postidentification feedback. Not surprisingly, evaluators were not perfectly able to make these discriminations. Ideally, testimony evaluators would believe 100% of accurate witnesses and 0% of mistaken witnesses. Nevertheless, in the absence of any feedback influence on the witnesses, evaluators were twice as likely to believe accurate witnesses as they were to believe mistaken witnesses. Moreover, in the absence of feedback, there was no evidence of overbelief as indicated by the fact that false positives (believing a mistaken witness) and false negatives (disbelieving an accurate witness) were not significantly different (36% and 30%, respectively).

Before discussing the moderating role of postidentification feedback, it is important that we note a caveat regarding evaluators' abilities to discriminate between accurate and mistaken identification testimony in the control (no feedback) conditions. Specifically, we caution readers against presuming that this level of

discrimination in the no-feedback conditions is what one would expect in actual cases. Unlike identification procedures that are recommended in actual cases, we used the standard postidentification feedback paradigm in which witnesses were given biased instructions without a no-identification option, resulting in positive identification decisions from all witnesses. Accordingly, some unknown portion of witnesses who were shown a culprit-absent lineup would probably have chosen to identify no one if they had been given that option. Presumably, this subset of witnesses would have made their mistaken identifications with little conviction. That lack of conviction among some of the mistaken eyewitnesses might have come across during testimony, thereby making the task of correctly discounting those mistaken witnesses easier for the evaluators. This could have resulted in inflated discrimination scores in the no-feedback conditions.

As noted in the Methods section, obtaining identifications from all witnesses is the common procedure used in postidentification feedback studies because of its efficiency. However, we favored this procedure for an additional reason. Specifically, we wanted to make sure that the evaluators could perform reasonably well in making discriminations between accurate and mistaken testimony in the control conditions so as to have a good conditions for testing the moderating role of feedback. In any case, one consequence of the method we used to manipulate accuracy is that it might have made the discrimination task easier for the evaluators in the control conditions, which could serve to undermine external validity. But the same method was used in the feedback conditions as well, which speaks to the internal validity of the study. Moreover, the fact that a simple comment given to witnesses following their identifications was able to wipe out discrimination speaks to the potency of the observed effect.

When witnesses were under the influence of confirming feedback, the ability of evaluators to discriminate between accurate and mistaken testimony was totally eliminated. Mistaken eyewitnesses who had received feedback ultimately delivered testimony that was just as credible as the testimony of accurate eyewitnesses, as evidenced by equal belief rates in the feedback conditions (about 63%). Furthermore, confirming feedback led to overbelief, which was evidenced by the fact that false positives ballooned to 63% under conditions of feedback, whereas false negatives remained low (36%).

The pattern of results is broadly consistent with the cue-accessibility interpretation of the postidentification feedback effect, which posits that when internal cues are weak witnesses will rely on external cues (e.g., feedback) to infer retrospective accounts on such matters as how much attention they paid, how good their view was, and how certain they were (Charman et al., 2010; Neuschatz et al., 2007; Wells & Bradfield, 1998, 1999). The larger effect of postidentification feedback on mistaken than on accurate witnesses is presumed to result from accurate witnesses having stronger internal cues (recollection of a sense of recognition) than do mistaken witnesses (see Bradfield et al., 2002; Charman & Wells, 2012).

This research also provides a unique assessment of the extent to which the traditional self-report measures of the feedback effect serve as adequate proxies for eyewitness testimony. Although the traditional measures have been used in almost all of the extant research of the postidentification feedback effect, the results of this study suggest that the measures might not fully capture important aspects of feedback's influences on eyewitness testimony. Specifically, after controlling for witnesses' self-reports on the standard measures of the postidentification feedback effect, the accuracy-by-feedback interaction effect on evaluators' belief judgments remained statistically significant, indicating that the self-report measures did not fully account for the effects of feedback on evaluators' belief judgments. A different result emerged, however, when we examined evaluators' ratings of the witnesses' narrative event descriptions. Ratings of the narrative descriptions not only showed the same interaction pattern as the belief data, but when we statistically controlled for the evaluators' ratings of the narrative descriptions, the interaction effect in the belief data was eliminated. Previous postidentification feedback experiments have not measured characteristics of witnesses' narrative descriptions. The current results suggest that understanding how postidentification feedback affects witnesses' narrative event descriptions could be important in understanding how feedback effects translate into testimony believability.

To date, more than 220 innocent individuals in the United States have been exonerated through DNA testing after having been wrongly convicted in cases involving mistaken eyewitness-identification testimony (*innocenceproject.org*, 2013). Unfortunately, only a small fraction of eyewitness-identification cases have biologically rich DNA trace evidence that can test the claims of an eyewitness. Hence, courts remain very dependent on the presumption that triers-of-fact can discriminate between accurate and mistaken eyewitness-identification testimony. The current results show that confirming postidentification feedback complicates the burden on triers-of-fact to perform their duty. The abysmal performance of testimony evaluators in the confirming feedback conditions was not the fault of the triers-of-fact; it was a result of feedback given to the witnesses.

Although our manipulation of confirming feedback followed the standard postidentification feedback paradigm practice of using an experimentally scripted response from the lineup administrator, it is likely that other forms of feedback would have similar effects on eyewitness testimony. For example, research shows that lineup administrators tend to naturally leak confirming feedback when they learn that the witness identified the person who they believe is the suspect (Garrioch & Brimacombe, 2001). Garrioch and Brimacombe misled lineup administrators to think that one or another member of a culprit-absent lineup was the suspect, whereas other administrators were kept blind. They found strong administrator expectancy-effects on witnesses' certainty that, when analyzed on videotapes, tended to be the result of cues involving intonation, gestures, facial expressions, head nodding, and eye contact. Importantly, 100% of the lineup administrators did not believe that they had provided any influence and 95% of the witnesses also reported that the lineup administrator provided no influence. This creates an especially difficult problem for prosecutors, judges, and juries who are dependent on testimony from the witness and the lineup administrator to discover possible feedback influences on the witness.

Partial solutions for the postidentification feedback problem have been proposed—in particular, the double-blind lineup in which the person administering the lineup does not know which person is the suspect and which are merely fillers (Wells, 1988). The double-blind lineup has generally been described as a mechanism for preventing lineup administrators from influencing wit-

nesses' identification decisions (e.g., see Greathouse & Kovera, 2009). But the double-blind lineup is also important for getting lineup administrators to make objective records of the witnesses' choices (see Rodriguez & Berry, 2012) and for securing "clean" statements (e.g., certainty statement) from witnesses immediately after any identification, before the witnesses' testimony-relevant judgments can be contaminated by feedback (Wells et al., 1998). Although witnesses will likely encounter confirming feedback later from prosecutors or infer confirmation simply from the fact that the identified person was indicted, empirical evidence indicates that getting witnesses to answer statements about certainty before receiving feedback has at least a partial prophylactic effect against later feedback inflation effects (e.g., Quinlivan et al., 2009; Wells & Bradfield, 1999). Moreover, if inflation does occur before trial, the prefeedback witness statements would then be discoverable records at trial, disclosed to the defense, and can be shown to the judge and jury. And, there is evidence that disclosing prefeedback witness statements can temper jurors' evaluations of eyewitness-identification testimony when certainty inflation has occurred (e.g., Bradfield & McQuiston, 2004; Douglass & Jones, 2013). Although witnesses might still be able to come up with an "explanation" if their certainty and other judgments inflated over time (see Jones, Williams, & Brewer, 2008), statements taken by a blind administrator at the time of identification (before any opportunity for feedback) represent important data for the triers-of-fact to do their job of sorting between accurate and mistaken identification testimony.

## References


Bradfield, A., & McQuiston, D. E. (2004). When does evidence of eyewitness confidence inflation affect judgments in a criminal trial? *Law and Human Behavior, 28,* 369–387. doi:10.1023/B:LAHU.0000039331.54147.ff

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. 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

Charman, S. D., Carlucci, M., Vallano, J., & Hyman-Gregory, A. (2010). The selective cues 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. (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

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

Desmarais, S. L., & Read, J. D. (2011). After 30 years, what do we know about what jurors know? A meta-analytic review of lay knowledge regarding eyewitness factors. *Law and Human Behavior, 35,* 200–210. doi:10.1007/s10979-010-9232-6

Douglass, A. B., & 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., Neuschatz, J. S., Imrich, J., & Wilkinson, M. (2010). Does post-identification feedback affect evaluations of eyewitness testimony and identification procedures? *Law and Human Behavior*. 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

Garrioch, L., & Brimacombe, C. A. E. (2001). Lineup administrators' expectations: Their impact on eyewitness confidence. *Law and Human Behavior, 25,* 299–315. doi:10.1023/A:1010750028643

Greathouse, S. M., & Kovera, M. B. (2009). Instruction bias and lineup presentation moderate the effects of administrator knowledge on eyewitness identification. *Law and Human Behavior, 33,* 70–82. doi:10.1007/s10979-008-9136-x

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

Innocenceproject.org, retrieved June 13, 2013.

Jones, E. E., Williams, K. D., & Brewer, N. (2008). "I had a confidence epiphany!": Obstacles to combating post-identification confidence inflation. *Law and Human Behavior, 32,* 164–176. doi:10.1007/s10979-007-9101-0

Kiesler, C. A. (1971). *The psychology of commitment: Experiments linking behavior to belief*. San Diego, CA: Academic Press.

MacLean, C. L., Brimacombe, C. A., 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

Neuschatz, J. S., Lawson, D. S., Fairless, A. H., Powers, R. A., Neuscahtz, 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. R., 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

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

Rodriguez, D. N., & Berry, M. A. (2012). The effect of line-up administrator blindness on the recording of eyewitness identification decisions. *Legal and Criminological Psychology*. doi:10.1111/j.2044-8333.2012.02058.x

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

Tetlock, P. E. (1983). Accountability and complexity of thought. *Journal of Personality and Social Psychology, 45,* 74–83. doi:10.1037/0022-3514.45.1.74

Tetlock, P. E., Skitka, L., & Boettger, R. (1989). Social and cognitive strategies for coping with accountability: Conformity, complexity, and bolstering. *Journal of Personality and Social Psychology, 57,* 632–640. doi:10.1037/0022-3514.57.4.632

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

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., Ferguson, T. J., & Lindsay, R. C. L. (1981). The tractability of eyewitness confidence and its implication for triers of fact. *Journal of Applied Psychology, 66,* 688–696. doi:10.1037/0021-9010.66.6.688

Wells, G. L., Lindsay, R. C. L., & Ferguson, T. J. (1979). Accuracy, confidence, and juror perceptions in eyewitness identification. *Journal of Applied Psychology, 64,* 440–448. doi:10.1037/0021-9010.64.4.440

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. doi:10.1023/A:1025750605807

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 March 19, 2013
Revision received August 27, 2013
Accepted September 12, 2013 ■