Palmer, M. A., Brewer, N., Weber, N., & Sauer, J. D. (2021). Eyewitness identifications of multiple culprits: Disconfirming feedback following one lineup decision impairs identification of another culprit. Psychology, Public Policy, and Law, 27(2), 170–183. https://doi.org/10.1037/law0000291

Eyewitness Identifications of Multiple Culprits: Disconfirming Feedback Following One Lineup Decision Impairs Identification of Another Culprit

Matthew A. Palmer
University of Tasmania

Neil Brewer
Flinders University

Nathan Weber
Flinders University

James D. Sauer
University of Tasmania

Matthew Palmer, School of Psychological Science, University of Tasmania, Hobart, Tasmania, Australia. Neil Brewer and Nathan Weber, College of Education, Psychology & Social Work, Flinders University, Adelaide, Australia. James Sauer, School of Psychological Science, University of Tasmania, Hobart, Tasmania, Australia.

We thank Nicole McCallum, Emily Palmer, and Cameron Palmer for help with data collection. This research was supported by Australian Research Council Grants DP1093210 to N. Brewer et al. and DP140103746 to M. Palmer et al. Parts of this research were presented at the biennial meeting of the Society for Applied Research in Memory and Cognition (2013).

Correspondence concerning this article should be addressed to Matthew Palmer, School of Psychological Science, University of Tasmania, Private Bag 30, Hobart, TAS 7001, Australia. E-mail: matthew.palmer@utas.edu.au

Abstract

Eyewitnesses to multiple-culprit crimes are often asked to try to identify the culprits from different lineups during a police investigation. In two experiments (total $N = 557$), we show that disconfirming feedback after an identification attempt for one culprit can impair identification performance on a subsequent lineup for a different culprit. In each experiment, witnesses viewed a simulated, two-culprit crime, followed by two police lineups: A culprit-absent lineup for one culprit, and either a culprit-present or culprit-absent lineup for the second culprit. Following the first lineup, witnesses received disconfirming feedback or no feedback. For witnesses who correctly rejected the first lineup, disconfirming feedback impaired identification performance on the subsequent lineup. For witnesses who incorrectly chose someone from the first lineup, disconfirming feedback impaired subsequent performance when the feedback unambiguously implied poor ability to identify the culprit in the first identification test (Experiment 2) but not when it could have been interpreted as implying poor criterion setting (Experiment 1). Across both experiments, disconfirming feedback also reduced the difference in confidence between correct and incorrect identifications. These results add to evidence that postidentification feedback can affect subsequent identification performance by influencing witnesses' beliefs about their ability to identify a culprit. Current policy recommendations state that postidentification feedback should be withheld from witnesses until confidence has been documented. These should be updated to recommend withholding feedback for longer if the witness may be asked to view additional lineups, and to ensure that lineup administrators are blind to the results of any previous lineups.

Keywords: eyewitness identification; multiple culprits; multiple lineups; postidentification feedback; compound signal detection

Disconfirming Feedback Impairs Subsequent Eyewitness Identification of a Different Culprit

Eyewitnesses to crimes are often asked to view a lineup and attempt to identify the perpetrator of the crime. Witnesses who make an identification decision often receive feedback about the accuracy of their decision. For example, a witness who selects the suspect from a lineup might be told so by the lineup administrator (“Well done – you picked the suspect!”). Postidentification feedback can be confirming (implying that the witness picked the suspect) or disconfirming (implying they did not). Feedback can be given explicitly to the witness or inferred by the witness (e.g., from body language), and can be provided by the lineup administrator or another witness (“I picked number three.”) Postidentification feedback affects witnesses’ ratings of confidence in their identification decision and retrospective reports of witnessing conditions. For example, witnesses who receive confirming feedback (relative to no feedback) express greater confidence in the accuracy of their decision, and report that while witnessing the crime they had a better view of the culprit, saw the culprit for longer, and were paying more attention to the culprit (Douglass & Steblay, 2006; Semmler et al., 2004; Steblay et al., 2014; Wells & Bradfield, 1998, 1999). These effects are well-established and have important implications for the legal system.

Recently, researchers have begun to investigate the extent to which postidentification feedback can influence performance on a subsequent identification attempt. Two studies have addressed this issue in situations in which a witness attempts to identify a single culprit from two different lineups (Palmer, Brewer, & Weber, 2010; Smalarz & Wells, 2014). In such cases, feedback following an initial lineup affects identification performance for a second identification decision.

We examined the effects of postidentification feedback in situations involving multiple culprits, whereby the witness receives feedback after one identification attempt and then attempts to identify a different culprit from a subsequent lineup. This is an important issue because (1) such cases occur quite frequently, (2) as we will explain, it cannot be assumed that the effects of feedback found in multiple lineups for a single culprit will translate to a scenario involving different culprits, and (3) the extent to which such effects occur has implications for procedures for collecting eyewitness evidence in multiple-culprit crimes.

Effects of Feedback on Subsequent Identification Performance

In various situations, witnesses are asked to view more than one lineup. In some cases, a witness might view multiple lineups in an attempt to identify a single culprit. For example, the police might arrest an initial suspect and place that person in a lineup. If new evidence then emerges that exonerates the original suspect and implicates a second suspect, the police might place the second suspect in a lineup for the witness to view. In other cases, a witness may view multiple lineups for different culprits. For example, if a crime involves more than one culprit (e.g., a bank robber and a getaway driver), the witness might be asked to view separate lineups: in this case, one lineup that contains a person suspected of being the robber, and another lineup that contains a person suspected of being the driver. Such situations occur frequently (Hobson & Wilcock, 2011). In one archival study of field data from the UK, 61% of lineup identification tests (512 of 833) were from cases that involved multiple culprits, and witnesses viewed multiple lineups for different culprits in many of these cases (Horry, Halford, Brewer, Milne, & Bull, 2014).

Researchers have recently found that postidentification feedback can affect performance on a subsequent identification test involving the same culprit. Palmer et al. (2010) had participants attempt to identify a culprit from two lineups: an initial lineup that was culprit-absent (i.e., did not contain the culprit) and a second lineup that was either culprit-absent or culprit-present (i.e., did contain the culprit). Participants were randomly allocated to receive true feedback or no feedback after the first lineup. Participants who chose from the initial culprit-absent lineup (*initial choosers*) received true disconfirming feedback that their first decision was incorrect. Participants who rejected the initial culprit-absent lineup (*initial non-choosers*) received true confirming feedback that their first decision was correct. Feedback influenced identification performance on the second test: Compared to no feedback, true confirming feedback increased accuracy for initial non-choosers, and true disconfirming feedback reduced accuracy for initial choosers.

Palmer et al. (2010) proposed that the underlying mechanism involves feedback affecting participants' perceptions of their ability to identify a culprit from an upcoming lineup. In a separate experiment, Palmer et al. found that true feedback after an initial lineup affected participants' expectations of making a correct response to an upcoming second lineup: True confirming feedback increased expectations of making a correct decision, and true disconfirming feedback reduced expectations. Palmer et al. suggested that changes in expectations of accuracy could have caused witnesses to alter the way that they used memorial information during the identification decision (cf. Lane, Roussel, Villa, & Morita, 2007). Disconfirming feedback—which reduces expectations of accuracy—could prompt witnesses to evaluate memorial evidence less stringently during a subsequent identification decision by, for example, placing less weight on the recall of vivid, distinctive information (Gallo, Weiss, & Schacter, 2004; Schacter, Israel, & Racine, 1999). Conversely, confirming feedback—which increases expectations of accuracy—could prompt witnesses to evaluate memorial evidence more stringently during a subsequent identification decision.

Smalarz and Wells (2014) had witnesses attempt to identify a single culprit from multiple lineups. Participants viewed and responded to a culprit-absent lineup, after which some participants received false confirming feedback for their decision. Participants were then shown a culprit-present lineup for the same culprit. The second lineup contained no members of the initial lineup. Participants who received false confirming feedback after their first lineup response performed worse on the second identification test than did participants who received no feedback. Smalarz and Wells suggested that false confirming feedback might alter the witness's memory of the culprit, such that the witness's memory of the culprit becomes more like—or perhaps is replaced by—the chosen lineup member. Thus, when making a future identification response, the witness would compare lineup members to the newer (incorrect) image, rather than the original image of the culprit. In turn, identification accuracy would suffer.

The account proposed by Smalarz and Wells is a plausible mechanism by which false feedback might impair performance on a subsequent attempt to identify that same culprit. However, it is not relevant in cases where a witness attempts to identify a different culprit from a second lineup. Because there is little reason to expect that false confirming feedback will alter a witnesses' memory for a different culprit, there is little reason to expect that false confirming feedback after an initial wrongful identification will impair a subsequent attempt to identify a different culprit.

Could Feedback Affect Identification of a Different Culprit?

If postidentification feedback was to influence subsequent attempts to identify a different culprit, how might this occur? The feedback mechanisms proposed by Palmer et al. (2010) and

Smalarz and Wells (2014) are not mutually exclusive and do not necessarily apply to the same situations. However, as noted above, the mechanism proposed by Smalarz and Wells only applies to situations where witnesses view multiple lineups for a single culprit. If feedback influences a witness's memory for one particular culprit, it will affect subsequent attempts to identify that same culprit, but not a different culprit.

In contrast, the mechanism proposed by Palmer et al. (2010) could result in feedback affecting a subsequent attempt to identify a different culprit if feedback affects perceptions of one's ability not just to identify the specific culprit from the first lineup, but another culprit. The broader literature on postidentification feedback and confidence inflation provides some evidence in support of this notion. Numerous studies have shown that postidentification feedback affects not only witnesses' confidence in that specific decision, but also—and crucially—witnesses' perceptions of factors that are not specific to one culprit, such as self-reports of their ability to recognize the faces of strangers (e.g., Bradfield, Wells, & Olson, 2002; Wells & Bradfield, 1998, 1999). These findings suggest that feedback might affect witnesses' perceptions of their ability to recognize not only the person they were trying to identify from an initial lineup (the one they received feedback for) but also their ability to recognize other unfamiliar faces. If this leads to changes in the way individuals monitor their memory or place their decision criterion, then feedback could affect accuracy on a subsequent attempt to identify a different person.

However, even though feedback seems to affect subsequent attempts to identify the same culprit, it cannot be taken for granted that feedback will influence subsequent identification of a different culprit. Performance on a memory task is most strongly related to beliefs about current memory ability in relation to that specific task; performance is less strongly related to beliefs about memory ability in the same broad domain, or beliefs about memory ability in general (e.g., Hertzog, Park, Morrell, & Martin, 2000; Beaudoin & Desrichard, 2011). For example, performance for remembering a specific list of words is more strongly related to beliefs about current memory ability for that list of words, compared to beliefs about memory ability for words in general, or about general memory ability. This is, in part, because people's beliefs about their memory ability can vary between different memory tasks (e.g., Crook & Larabee, 1990; Hertzog, 2002). Moreover, for a particular memory task, such as attempting an identification decision, beliefs about memory ability include not only relatively stable beliefs about one's general ability to perform that type of task (*task-specific memory self-efficacy*) but also temporary beliefs about one's ability to perform a specific example of the task at the present time (termed *concurrent memory self-efficacy*; Hertzog & Dixon, 1994). These two types of beliefs do not necessarily co-vary. A person who thinks that they generally have a poor memory for names, and who recently learned they misremembered a name, might still think they are highly likely to remember a specific name in a particular situation (e.g., because they formed a strong association between that name and a relative of theirs with the same name). Similarly, consider a witness who thinks they are generally poor at recognizing faces, and recently learned they made an incorrect identification decision from a lineup. That witness might still be confident in their ability to identify a different culprit from another lineup, for example, if they have a vivid memory of the second culprit or can recall a specific, distinctive feature.

From this viewpoint, feedback following an identification decision for one culprit might affect perceptions of ability to identify that specific culprit, but not other culprits. If so, postidentification feedback will have little effect on a subsequent attempt to identify a different culprit. The focus of this research was to distinguish between these possibilities.

Effects of Feedback on Confidence for Correct and Incorrect Identifications

Although not our main focus, we also examined the effects of disconfirming feedback on confidence for correct and incorrect identifications. Confidence plays a crucial role in evaluating the accuracy of identification decisions. Mistaken identification has been described as the leading cause of wrongful imprisonment in cases of DNA evidence-based exonerations (Innocence Project, 2016), and researchers have dedicated much effort to trying to find reliable markers that aid in evaluating the accuracy of identification decisions. Confidence is by far the most well-established of such markers. Although researchers are still working towards a comprehensive understanding of the utility of confidence for evaluating individual identification decisions (Sauer, Palmer, & Brewer, 2019), a considerable body of evidence shows that correct identification decisions are made with higher confidence than incorrect ones for decisions where the witness identifies someone, but not where the witness rejects the lineup (e.g., Brewer & Wells, 2006; Palmer, Brewer, Weber, & Nagesh, 2013; Sauer, Brewer, Zweck, & Weber, 2010; Sporer, Penrod, Read, & Cutler, 1995; Wixted & Wells, 2017).

The relationship between confidence and accuracy is ideally assessed via the use of more sophisticated analysis of calibration (e.g., Brewer & Wells, 2006) or the confidence-accuracy-characteristic (CAC; Mickes, 2015). However, we took the more coarse-grained approach of comparing mean confidence for correct versus incorrect identification decisions for two reasons. First, our planned sample size was based on our primary research objective of investigating the effects of feedback on identification performance on a subsequent lineup. To include calibration or CAC analysis would require a much larger sample. Second, comparing confidence for correct and incorrect decisions still provides useful information about the utility of confidence for discriminating correct from incorrect identification responses.

How might feedback affect confidence for correct and incorrect identifications? If disconfirming feedback following an initial lineup reduces discriminability for a second lineup, then it could also reduce the difference in confidence between correct and incorrect identifications from the second lineup. According to signal detection theory (SDT) models, decision confidence depends on the amount by which strength of evidence exceeds the decision criterion (e.g., Stretch & Wixted, 1998). When discriminability (the ability to distinguish targets from fillers) is poorer, the difference in confidence between correct and incorrect picks will become smaller. This is because—all else being equal—poorer discriminability means that a smaller proportion of the distribution of strength of evidence for targets will exceed the decision criterion, and a larger proportion of the distribution for fillers will exceed the decision criterion. Thus, when a target is correctly identified, the amount by which strength of evidence exceeds the decision criterion will, on average, be smaller under poor discriminability than good discriminability (producing lower confidence in correct decisions). Conversely, when an incorrect identification is made, the amount by which strength of evidence exceeds the decision criterion will, on average, be greater under poor discriminability than good discriminability (producing higher confidence in incorrect decisions). As a result, the difference in mean confidence should be smaller under poor discriminability than good discriminability.

We conducted two experiments to investigate the effects of disconfirming feedback following an initial lineup on identification of a different culprit from a second lineup. We report the effects on identification responses for both experiments first, then effects on confidence for correct and incorrect decisions.

Experiment 1

Experiment 1 provided an initial test of the effects of feedback on identification of a different culprit. Mock witnesses viewed a video of a simulated crime committed by two culprits. Witnesses then viewed two lineups and attempted to identify one culprit from the initial lineup, and the other culprit from the subsequent lineup. The initial lineup that each witness viewed was always culprit-absent (i.e., did not contain the person the witness was looking for). The second lineup was either culprit-absent or culprit-present (contained the person the witness was looking for). Following the first lineup—and prior to viewing the second lineup—witnesses received disconfirming feedback or no feedback about their initial identification response. A control group of witnesses viewed only a single culprit-absent or culprit-present lineup; they did not view the initial culprit-absent lineup.

Our aim was to conduct an initial test of whether feedback could influence a subsequent attempt to identify a different person. We acknowledge that a thorough investigation of this issue would require the use of confirming and disconfirming feedback, and culprit-present (as well as culprit-absent) initial lineups. However, logistical constraints restricted us to including only one type of feedback, and we had two reasons for focusing on disconfirming feedback rather than confirming. First, disconfirming feedback provides a stringent test of the basic notion that feedback following one identification decisions might affect subsequent identification of a different culprit. The effects of disconfirming feedback on confidence tend to be weaker than those of confirming feedback (e.g., Charman, Carlucci, Vallano, & Hyman Gregory, 2010; Steblay et al., 2014). Thus, to the extent that the same patterns hold in the current context, disconfirming feedback represents the more conservative test of whether feedback can influence subsequent attempts to identify a different culprit. Second, although confirming feedback might occur more frequently in real police investigations, it is realistic to expect that some witnesses in police investigations receive disconfirming feedback for not picking the suspect. Explicit disconfirming statements might be rare ("the person you picked was actually not the suspect we are investigating") —although how rare is unknown—but more subtle expressions of disconfirming feedback may be more realistic (e.g., if the lineup administrator's body language conveys disappointment with the identification response; Garrioch & Brimacombe, 2001). However, our goal here was not to replicate the most likely form of feedback in actual cases, but to test the basic idea that feedback might affect a subsequent attempt to identify a different culprit. Based on the results of Palmer et al. (2010), we expected that discriminability on a second identification test would be poorer for witnesses who received disconfirming feedback (rather than no feedback) for an initial lineup.

Method

Participants and design. Participants were 330 paid university students (196 female; aged 17-61 years, $M = 23.06$, $SD = 6.75$). A further 33 participants were excluded: 28 who expressed suspicion about the feedback manipulation and five who personally knew one of the targets from the stimulus video. Data from these participants were not analyzed. Participants were randomly allocated to one condition of a 3 (Feedback: disconfirming feedback; no feedback; single-lineup control group) × 2 (Lineup type: culprit-present; culprit-absent) between-subjects design. For data analysis, the disconfirming feedback and no feedback conditions were further split into initial choosers and non-choosers, based on responses to the first lineup.

We did not have a specific stopping rule for data collection but recruited as many participants as possible within the available funding window, aiming for approximately 30 participants per cell. The balance of participants across cells in Experiment 1 was complicated by two factors: (1) We could not anticipate the split between witnesses who chose from or rejected the initial lineup, and (2) participants were randomly allocated to receive feedback or no feedback prior to their response to the initial lineup. Thus, we could not ensure equal numbers of participants across different cells for analysis.

Procedure and materials. This research was approved by the relevant institutional human research ethics committee (Project 4984). Participants completed the experiment alone in a quiet laboratory cubicle. A computer was used to administer all instructions and record participants' responses. All participants viewed a mock-crime video, completed a 5-min distractor task, and attempted to identify either one or two culprits who had appeared in the video.

Participants in the single-lineup control group attempted to identify one culprit from a 6-person lineup that contained the culprit and five fillers (a culprit-present lineup) or six fillers (a culprit-absent lineup). To ensure that our results were not idiosyncratic to particular materials, we used three sets of stimuli that have been used in previous eyewitness identification studies (Brewer & Wells, 2006; Charman & Wells, 2007; Palmer et al., 2010). Each stimulus set comprised a video of a different mock-crime (three different instances of theft) and a lineup for each of two target persons from that video. For each stimulus set, lineup fillers were chosen on the basis that they matched the description of the culprit (details can be found in the citations above). One stimulus set was used for each participant. Counterbalancing ensured that each of the six target persons served equally often as the focus of the first and second lineups. As intended, the stimuli produced variation in identification patterns without approaching floor or ceiling performance: Collapsing across target-present and -absent lineups (on the second identification test), accuracy ranged from 34% to 63% across the six targets; and choosing rates from 29% to 73%. For all other analyses, data were collapsed across the three stimulus sets and six individual targets.

Prior to viewing the lineup, the instructions specified which person the participant was to look for (e.g., “In the video there was a male wearing a red shirt and glasses. We would like you to try to identify him from a lineup.”) Participants were informed that the person they were looking for may or may not be in the lineup, and that they should respond by clicking the face of the person (if they are in the lineup) or the “not present” button. These instructions, along with a specification of the culprit to be identified, preceded each lineup in every condition. After making a response to a lineup, participants rated their confidence that they had made a correct decision on a scale ranging from 0% to 100% in decile increments.

Participants in the multiple-lineups groups attempted to identify two culprits from the video from separate 6-person lineups. The first lineup was always culprit-absent; the second was either culprit-present or culprit-absent. Participants in the multiple-lineups groups were randomly allocated to receive disconfirming feedback (“The response you made to the previous lineup was incorrect”) or no feedback following the first identification test. There were no other manipulations or experimental conditions.

Signal detection analyses. In assessing the effects of feedback on identification performance, we wanted to tease apart differences in witnesses’ ability to distinguish the culprit from other lineup members (discriminability) from effects on witnesses’ tendency to pick from or reject a lineup (response bias). This is crucial for understanding effects on eyewitness identification performance (e.g., Meissner et al., 2005; Mickes et al., 2012; Palmer & Brewer, 2012). To this end, we calculated estimates of d' (to index discriminability) and c (to index response bias) using an integration model of Signal Detection Theory-Compound Decision

(SDT-CD; Duncan, 2006). SDT-CD has been applied to lineup data in previous research (e.g., Palmer & Brewer, 2012; Smith, Wells, Smalarz, & Lampinen, 2018; Vitriol, Appleby, & Borgida, 2019) and is one of several approaches for estimating discriminability and response bias for eyewitness identification decisions (e.g., Lee & Penrod, 2019; Mickes, Flowe, & Wixted, 2012; Smith, Yang, & Wells, 2020; Wixted, Vul, Mickes, & Wilson, 2018).

SDT-CD is a signal detection model specifically designed for compound decisions, such as eyewitness identification decisions, that involve a *detection* component and an *identification* component. The detection component of the decision requires the witness to make a yes/no recognition decision about whether the target is somewhere in the lineup, regardless of which lineup member it is (i.e., is the target person in the lineup somewhere?). The identification component is an *m*- alternative forced choice recognition task, where *m* is the number of lineup members (i.e., assuming the target is in the lineup, which person is it?). Unlike models for simple recognition decisions (e.g., Gronlund, 2004; Meissner, Tredoux, Parker, & MacLin, 2005), SDT-CD provides an estimate of discriminability that reflects the compound nature of eyewitness identification decisions (for further details, see Duncan, 2006; Palmer & Brewer, 2012; Palmer et al., 2010). Two versions of SDT-CD, termed integration and independent observations models, vary in the decision rule for the detection component (Duncan, 2006; Wixted et al., 2018); we fit both to our data.

To estimate the variance for d' and c for each condition, we used a modified jackknife procedure to estimate the standard error for each value (Mosteller & Tukey, 1968; for examples, see Koriat, Lichtenstein, & Fischhoff, 1980; Weber & Brewer, 2006). For each condition, we calculated d' and c as many times as there were participants in that condition, each time omitting a different participant from the calculation. Thus, in a condition with 100 participants, d' and c

would be calculated 100 times, each time based on responses from 99 participants. This produced a distribution of *d′* and *c* pseudo-values for each condition, which were used to calculate a jackknife estimate of the standard error for *d′* and *c* in each condition.

Comparisons between feedback conditions were assessed via 95% inferential confidence intervals (95% ICIs). Following the procedure described in Tryon (2001), jackknife standard errors were used to compute 95% ICIs. Non-overlapping 95% ICIs indicate a difference at the alpha = .05 level for that particular comparison. ICIs are typically calculated for specific pairwise comparisons; however, we used an average *E* parameter, which enables comparisons between any pair of *d′* values or *c* values (see Tryon, 2001). Thus, although our focus was on comparing the feedback versus no feedback conditions within initial choosers and non-choosers, the 95% ICIs can also be used to compare, for example, the single lineup condition with other conditions.

Results

Of participants who viewed two lineups, 160 (62%) rejected the first lineup (initial non-choosers) and 98 (38%) chose a filler from the first lineup (initial choosers). Table 1 shows identification responses for each condition, and the single-lineup control group.

[Table 1 about here]

Model fitting. In using SDT-CD to estimate discriminability and response bias, observed rates of different identification responses (false identifications from target-absent lineups; correct identifications and filler identifications from target-present lineups) are compared to model-generated response rates. Note that the false identifications include all positive responses to target-absent lineups. The purpose of estimating *d′* and *c* is to describe the decision-making performance of witnesses, rather than assess the forensic value of identification decisions. For describing memory performance, all incorrect positive responses to target-absent lineup must be counted as false-positive errors; for assessing the forensic value of decisions, it is often useful to omit known false positives (Palmer & Brewer, 2012).

The model-generated rates are based on the notion that a single value of *d′* and a single value of *c* underpin both the detection and identification components of the lineup response. For each condition, we identified the combination of model-generated response probabilities that best fits the observed response rates via a series of log-likelihood ratio tests. These assess model fit for (a) correct identification from target-present lineups, (b) filler identifications from target-present lineups, and (c) filler identifications from target-absent lineups. The log-likelihood ratio tests yield a single G_{total} statistic for each condition, with smaller values indicating better fit.

We fit two versions of SDT-CD: an integration model and an independent observations model (Duncan, 2006). Although both models provided an adequate fit to the data, the integration model yielded smaller G_{total} values, indicating better fit to the data. Accordingly, we report estimates of *d′* and *c* based on the integration model. Estimates based on the independent observation model appear in Supplemental Material (Table S1) along with fit statistics for both models, including observed and model-generated response rates for each condition (Tables S2 and S3).

Effects of feedback on *d′* and *c*. Table 2 shows estimated *d′* and *c* values. Better ability to distinguish culprits from fillers is indicated by higher, positive values of *d′*. For *c*, positive values indicate a tendency to respond too conservatively (too many “not present” responses) and negative values indicate a tendency to respond too leniently. Following Palmer et al. (2010), we focused on the effects of postidentification feedback separately for initial choosers and initial non-choosers.

Feedback was associated with poorer discriminability for initial non-choosers, indicated by a lower *d′* for those who received disconfirming feedback than no feedback. However, for initial choosers, there was little difference in *d′* between those who received feedback and those who received no feedback.

In terms of response bias, for initial non-choosers there was little difference in *c* values between those who received feedback and no feedback. For initial choosers, overlapping ICIs indicated no statistically significant difference in bias, although the estimated *c* values hint at more conservative responding in the feedback condition (c = -0.04) than the no feedback condition (c = -0.52).

[Table 2 about here]

There was no evidence of a meaningful difference in discriminability or response bias between the single lineup control condition and any of the other conditions.

Discussion

The results for initial non-choosers show that disconfirming feedback (indicating that an initial lineup rejection was incorrect) was associated with poorer discriminability on a subsequent attempt to identify a different culprit. This demonstrates that the effects of feedback for a memory decision involving one stimulus can transfer to a decision involving a different stimulus.

However, for initial choosers, feedback had little effect on discriminability for the second identification decision. One explanation for these results is that the specific wording of the feedback given to participants may have been ambiguous for initial choosers. Participants who received feedback were informed that “the response you made to the previous lineup was incorrect”. Initial choosers may have interpreted this feedback in one of two ways, and these two interpretations would be expected to produce different effects of feedback on subsequent identification performance. As intended by the researchers, initial choosers may have attributed their previous incorrect response to poor ability at the identification task (e.g., due to having a poor memory for the culprit). This would be expected to result in reduced discriminability on a subsequent lineup, given that the effects of feedback operate via differences in perceptions of ability (as suggested by Palmer et al., 2010). Alternatively, initial choosers may have attributed their earlier incorrect identification to having been too willing to pick someone from the lineup. That is, initial choosers may have interpreted the feedback as implying that the culprit was not actually in the first lineup, and that their identification performance would have been better if they had adopted a more conservative approach to choosing from the lineup. Rather than resulting in reduced discriminability on a second identification test, this type of attribution would be most likely to produce more conservative responding on a second test. Thus, it could be that initial choosers in Experiment 1 attributed their earlier incorrect response to being too lenient in their responding, rather than poor ability at the identification task, and that this attribution produced more conservative responding for the second lineup. The response bias data from Experiment 1 offer some (weak) support for this notion: For initial choosers, disconfirming feedback—that an initial lineup pick was incorrect—was associated with a trend toward more conservative responding on the second lineup.

Experiment 2

In Experiment 2, we modified the wording of the disconfirming feedback statement to unambiguously imply that performance on the initial test was not simply due to overly lenient responding. To achieve this, we used feedback stating that the culprit had been present in the earlier lineup, but the witness did not pick them. It cannot be inferred from this feedback that more conservative responding would have improved identification accuracy, because a "not present" response to the earlier lineup would also have been incorrect.

Because we were specifically interested in the effects of the modified feedback on the subsequent identification performance of initial choosers in Experiment 2, we omitted the single lineup control condition and changed the initial identification task so that all participants were required to make a positive identification from the initial (target-absent) lineup. Hence, all participants in Experiment 2 were "initial choosers", with some receiving disconfirming feedback and others no feedback. This modification was made to reduce the number of participants required in order to address the research question. Although forcing witnesses to choose someone from the initial lineup reduces ecological validity, we reasoned that this design was unlikely to artificially inflate any feedback effects. If anything, the effects of feedback might be weaker for a forced decision than a voluntary one, given that participants would perhaps be less likely to take to take to heart feedback for a forced decision than a voluntary one. Thus, Experiment 2 likely represents a conservative test of the effects of disconfirming feedback.

Method

Participants. Participants comprised 227 paid undergraduate students (135 female) aged 17 to 52 years (*M* = 20.52 years, *SD* = 4.46 years). A further 12 participants were excluded on the basis that they expressed suspicion about the feedback manipulation; their data were not analyzed. As in Experiment 1, we recruited as many participants as possible within the available funding window, aiming for approximately 30 per cell.

Materials and procedure. These were the same as Study 1 except for the following modifications. The wording of the feedback was changed such that, following a response to the initial lineup, participants in the feedback condition were told "The response you made was incorrect. The [male wearing a red shirt and glasses] was in the lineup but you did not choose his photo.” Prior to viewing the initial lineup, participants were instructed: “In the video there was a [male wearing a red shirt and glasses]. We would like you to try to identify him from a lineup. Please click the face of the person you are looking for in the lineup.” There was no *not present* response option provided. For the second lineup, participants were informed that the person they were looking for may or may not be in the lineup, and asked to click the face of the person (if they are in the lineup) or the “not present” button. As in Experiment 1, there were no other manipulations or experimental conditions.

As in Experiment 1, we used an integration model of SDT-CD to estimate *d′* and *c*. Estimates of variance for *d′* and *c*, and associated 95% ICIs were calculated as in Experiment 1.

Results

Table 3 shows identification responses for each feedback condition in Experiment 2.

[Table 3 about here]

Effects of feedback on d′ and c. Table 2 (lower panel) shows estimates of *d′* and *c* for each condition along with standard errors and 95% ICIs. Feedback was associated with significantly poorer discriminability, evidenced by a lower *d′* value for the disconfirming feedback condition than the no feedback condition. Feedback had minimal effect on response bias.

It should also be noted that participants in Experiment 2 chose more often from lineups than those in Experiment 1 (*c* values ranging from – 0.04 to – 0.52). The forced-choice identification decision participants made at the initial lineup in Experiment 2 may have prompted participants to adopt a more liberal response criterion for the second identification test.

Discussion

In Experiment 2, compared to no feedback, disconfirming feedback (implying an incorrect pick from an initial lineup) was associated with poorer discriminability on a second identification test involving a different culprit. This is consistent with the notion that disconfirming feedback impairs identification for a subsequent decision when the feedback implies poor ability at the task, rather than inappropriate criterion setting. This result also demonstrates that postidentification feedback can influence subsequent identification performance for a different culprit, not only when the initial response was a lineup rejection (as in Experiment 1) but also when the initial response was an incorrect pick (Experiment 2). The implications of these results will be further considered in the General Discussion.

Effects of Disconfirming Feedback on Confidence for Correct and Incorrect Identifications

We compared identification confidence for correct and incorrect responses to the second lineup for decisions in which the witness picked someone from the second lineup (i.e., we excluded responses where the witness rejected the second lineup). As noted in the Introduction, although this issue is important, it was not our focus. We took this analytical approach because it provides useful information about the utility of confidence for distinguishing correct from incorrect identification decisions in the absence of calibration or CAC analyses (which were not feasible given our sample size per condition).

Figure 1 shows estimates of effect size (Cohen's *d*) for the difference in confidence between correct and incorrect identification decisions in each feedback condition in Experiments 1 (top panel) and 2 (bottom panel). Consistent with predictions based on SDT, Cohen's *d* values were smaller in the disconfirming feedback conditions than the no feedback conditions in Experiments 1 and 2. We used Exploratory Software for Confidence Intervals (ESCI; Cumming, 2012; 2014) with a random effects model to calculate two unbiased, meta-analytic effect size estimates—one for the no feedback conditions and one for the disconfirming feedback conditions—across both experiments. For witnesses who received no feedback, the difference in confidence between correct and incorrect decisions represented a moderate-large effect that was significantly different to zero, d = 0.71, 95% CI [0.38, 1.04], t = 4.23, p < .001. In contrast, for witnesses who received disconfirming feedback, the difference in confidence between correct and incorrect decisions was small and not significantly different to zero, d = 0.27 [-0.12, 0.66], t = 1.37, p = .17. These results suggest that disconfirming feedback reduced the difference in confidence between correct and incorrect identification decisions for a subsequent lineup. We expand on this point in the General Discussion.

[Figure 1 about here]

General Discussion

These results extend previous work showing that feedback following one identification decision can influence not only confidence in that decision but also identification performance on a subsequent lineup. Previous work showed this applied when the second identification attempt concerned the same culprit as the first attempt (Palmer et al., 2010; Smalarz & Wells, 2014). Our results extend this to provide the first demonstration that disconfirming feedback following a lineup response for one culprit can impair subsequent identification of a different culprit. It might seem unintuitive that feedback could worsen performance on a subsequent memory test, especially given that feedback is widely considered beneficial for learning and long-term memory (e.g., Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Roediger & Butler, 2011). However, numerous studies have shown that decision making can be impaired by feedback. This can occur, for example, when feedback is misleading (e.g., Lane at al., 2007) or in situations where providing the outcome of previous decisions does not support the improvement of future decisions (e.g., Brehmer, 1980; Hammond, Summers, & Deane, 1973). The results of Palmer et al. (2010) and Smalarz and Wells (2014) align with this second category of studies.

It might also seem unintuitive that disconfirming feedback did not prompt witnesses to adjust their response criterion (e.g., by setting a more liberal criterion after receiving feedback that an initial lineup rejection was incorrect). However, there is precedent in the literature that witnesses do not necessarily adjust their response criterion in a logical way when making multiple identification attempts. For example, consider a variation on the “blank lineup” procedure (Wells, 1984), in which witnesses are asked to identify one culprit and are told in advance that they will be viewing two lineups with no faces repeated across the lineups. Logically, if a witness chooses someone from the first lineup, they should be more likely to reject the second lineup based on commitment effects (e.g., Gorenstein & Ellsworth, 1980; Palmer et al., 2012). However, a recent study found minimal evidence of such logical shifts in decision criterion (Kucina, Sauer, Holt, Brewer, & Palmer, 2020). Such results—along with the present results and those of Palmer et al. (2010)—suggest that more research is required to understand the factors that shape criterion setting in multiple lineup scenarios.

In terms of the mechanism underpinning the effects of feedback on subsequent lineup responses, Palmer et al. (2010) showed that disconfirming feedback caused witnesses to lower their expectations about their ability to perform the identification task at hand. Palmer et al. proposed that this, in turn, prompted witnesses to (consciously or not) evaluate memorial evidence less stringently during a subsequent identification attempt, resulting in poorer discriminability. Our results are consistent with Palmer et al.’s account. Particularly noteworthy is the finding that, in our studies, disconfirming feedback impaired witnesses' performance on a subsequent lineup only when the feedback unambiguously implied poor ability at the identification task (i.e., for initial non-choosers in Experiment 1 and initial choosers in Experiment 2, but not for initial choosers in Experiment 1).

The fact that feedback influenced performance on a subsequent lineup for a different target has implications for the account proposed by Palmer et al. (2010). If the effects of feedback extended only to a subsequent attempt to identify that same target, this would suggest that feedback shapes expectations about performance—and the way that memorial information is evaluated—on a specific example of one type of memory task (akin to *concurrent memory self-efficacy*; Hertzog & Dixon, 1994). However, this was not the case: feedback instead seems to shape expectations about performance—and evaluation of memorial information—on a memory task that is related to—but not identical to—the task for which feedback was received (akin to *task-specific memory self-efficacy*). This finding raises questions about how broadly feedback effects might extend. For instance, could postidentification feedback influence performance on a subsequent attempt to identify a culprit from a different event? Or on a different type of memory test, such as recalling details of the witnessed event? These are issues for future research.

Our results also showed that disconfirming feedback was associated with a smaller difference in confidence between correct and incorrect identifications, suggesting that disconfirming feedback impaired the utility of confidence for discriminating correct from incorrect decisions. This, in turn, would reduce the ability of police investigators to evaluate the accuracy of identification evidence, and may cause additional problems for eyewitness decision making, given that witnesses rely on confidence to decide whether to report a decision or respond *don't know* (Perfect & Weber, 2012; Weber & Perfect, 2012).

It is important to note that this effect of feedback on identification confidence is different to the well-established postidentification feedback effect, whereby feedback for a decision influences confidence in that same decision (Douglass & Steblay, 2006; Steblay, et al. 2014; Wells & Bradfield, 1998, 1999). Our results show that feedback for one decision influences an important property of the confidence-accuracy relationship for a different decision, one for which no feedback was received.

These results align with predictions derived from SDT that the difference in mean confidence between correct and incorrect decisions will be smaller when discriminability is poorer. This conclusion holds up under close inspection of the results for different groups. Recall that disconfirming feedback reduced discriminability in Experiment 2 (d' for no feedback: 1.56 vs. feedback: 0.90) and for initial non-choosers in Experiment 1 (no feedback: 1.92 vs. feedback: 0.97), but not for initial choosers in Experiment 1 (no feedback: 1.59 vs. feedback: 1.71). The pattern of confidence-accuracy data reflects this. Following disconfirming feedback, the difference in confidence between correct and incorrect decisions represented a small effect in Experiment 2 (d = 0.27) and a negligible effect for initial non-choosers in Experiment 1 (d = 0.03), but for initial choosers in Experiment 1 this difference remained large (d = 0.82). This pattern of results is consistent with the idea that differences in confidence-accuracy resolution accompanied differences in discriminability.

Implications for Policy

Our results show that disconfirming feedback after an identification attempt can impair discriminability on a subsequent attempt to identify a different culprit. Moreover, disconfirming feedback can also reduce the utility of confidence for evaluating the likely accuracy of a subsequent identification decision. Both these outcomes are detrimental for police investigations, and they point to clear implications for policies regarding the collection of eyewitness identification evidence.

Our results underscore the importance of withholding feedback from eyewitnesses who, as the investigation proceeds, may be asked to attempt another identification test. Current recommendations already state that postidentification feedback can distort identification confidence and, thus, confidence statements should be collected and documented immediately after an identification decision, before any feedback is given (National Research Council, 2014; Steblay et al., 2014; Wells et al., 2020). However, these recommendations do not speak to situations in which feedback is provided after the identification decision is made and confidence is documented. For example, a conscientious investigator who is familiar with current recommendations may conduct an identification test and document identification confidence immediately after the witness's decision, and then provide feedback to the witness. This type of feedback cannot distort the original identification decision or confidence in that decision. However, our results show that it can affect the accuracy of an identification decision for a subsequent lineup—and the utility of confidence for discriminating correct from incorrect decisions—even if the subsequent lineup involves a different culprit. This knowledge is unlikely to be intuitive to police investigators, even those familiar with current recommendations about feedback.

Our research demonstrates the need to update current recommendations so that postidentification feedback is withheld not only until confidence has been recorded but also until there is no longer a possibility that the witness may be asked to view another lineup. This adjustment is especially important because, for various reasons, witnesses often view multiple lineups during police investigations (Horry et al., 2015; Kucina et al., 2020; Palmer, Brewer, &

Weber, 2012; Smalarz et al., 2019; Steblay et al., 2013; Wells, 1984), and many of these cases involve multiple culprits (Hobson & Wilcock, 2011; Horry et al., 2014). Thus, there is considerable scope for feedback to increase the incidence of eyewitness misidentification in cases involving multiple culprits. Given the prominent role of misidentification in cases of wrongful conviction (Innocence Project, 2016), it is worth exploring policies to reduce the likelihood of identification error.

What procedures are needed to address this issue? One obvious part of the solution is to use double-blind testing, where the lineup administrator is blind to which lineup member is the suspect. The need for double-blind testing has been emphasized in many recommendations for policy and procedure (e.g., National Research Council, 2014; Wells et al., 2019; Wells, Small, Penrod, Malpass, Fulero, & Brimacombe, 1998); our research merely underscores its importance. Without double-blind testing, it may be very difficult indeed for investigators to avoid conveying some form of non-verbal feedback if the witness does not pick the suspect (e.g., an expression of disappointment). Ideally, witnesses would have no interaction with a lineup administrator until all identification tests have been completed. This can be achieved, for example, via computer-administered identification tests (Cutler, Daugherty, Babu, Hodges, & Van Wallendael, 2009).

However, even with double-blind testing, extra care must be taken to prevent witnesses from receiving feedback in cases involving multiple culprits. For example, if a witness is required to attend the police station to view lineups on separate occasions (e.g., if different suspects are arrested at different times), a lineup administrator might convey feedback about the first decision before the witness views the second lineup (e.g., “you did not pick the suspect last time but just try your best today”). Note that this could occur even if the administrator is blind to the identity of the suspect in the second lineup; the critical issue is whether the administrator is aware of the witnesses' earlier decision. Simply informing investigators that they should not provide feedback is unlikely to mitigate this problem, because administrators can convey feedback unintentionally (e.g., Greathouse & Kovera, 2009). A more effective approach is to ensure that the witness has no contact with an administrator who knows of their earlier lineup responses until all identification tests have been completed. In some cases, this can be done by conducting all lineups back-to-back in a single session. For example, in a case involving two culprits, police might apprehend two suspects and place them in two separate lineups for the witness to view in a single, computer-administered session.

In other cases, however, it will not be possible to prepare all lineups in advance. For example, investigators may have apprehended one of two suspects and wish to conduct a lineup for that person while still searching for the second suspect. In cases where the witness is likely to view additional lineups in the future, investigators could use a naïve administrator for each lineup who is blind not only to the identity of the suspect in that lineup, but also to previous identification decisions made by the witness. Investigators involved in such a case would quite rightfully be interested in the outcome of the first lineup; indeed, that outcome may well influence who they target as the second suspect. Thus, there would clearly be the possibility for interactions between these investigators and whoever is administering any subsequent lineup. Thus, there is considerable scope for leakage of information about previous lineups. Developing a foolproof way of dealing with this issue will likely be difficult, and rigid protocols will be needed to ensure that lineup administrators are blind to the result of previous lineups.

Limitations and Future Directions

Although our findings inform understanding of the effects of feedback in cases involving eyewitness identification of multiple culprits, it is important to acknowledge the limitations of this research and how these might help shape future research in this area. One limitation concerns the design of Experiment 2, in which the first lineup task was a forced-choice decision (i.e., witnesses had to pick someone) with biased instructions (i.e., no warning that the target person might not be present), whereas the second lineup included a “not present” option and unbiased instructions (a warning that the target person might not be present). This approach clearly lacks ecological validity and—as noted in the results section for Experiment 2—the forced-choice initial lineup task might have prompted participants to set a more lenient response criterion than they otherwise would have. However, it is unlikely that this magnified the effects of feedback in Experiment 2; if anything, the use of a forced-choice initial lineup might have dampened the effects of feedback, given that disconfirming feedback might not have been taken to heart by participants who would have correctly rejected the first lineup had they been allowed to do so. The results of Experiment 2 do, however, raise the broader issue of whether the effects of feedback might vary depending on propensity to choose from a lineup. For example, individual differences in response criterion (Kantner & Lindsay, 2012) might shape the extent to which post-identification feedback influences subsequent identification responses. The specific content of feedback in Experiment 2 also departs from what would likely occur in a police investigation: Rather than claiming that the culprit was in the lineup and had not been picked (as in Experiment 2), a police lineup administrator might be more likely to convey that the witness had not picked the suspect (without stating explicitly that the culprit was in the lineup). It is an empirical question whether such a difference in wording would alter the effects of feedback.

Another limitation is the use of disconfirming—but not confirming—feedback in both experiments. As outlined in the Introduction, we chose to use disconfirming feedback partly for pragmatic reasons (to keep the number of participants manageable) and because it provided a more stringent test of our hypotheses (given that post-identification feedback effects on confidence tend to be stronger for confirming than disconfirming feedback). Our results show that disconfirming feedback can impair discriminability on a subsequent lineup for a different culprit. In light of the results of Palmer et al. (2010), this suggests that confirming post-identification feedback might enhance discriminability on a subsequent lineup for a different culprit. However, we cannot take this for granted, and future research should examine this issue.

Additionally, it remains an empirical question as to what factors might moderate effects of feedback on subsequent identification attempts. For example, the effects of feedback might vary depending on strength of memory for the culprit. The effects of post-identification feedback on confidence tend to be stronger when internal, memory-based cues for confidence are weak, as might be expected when memory for the culprit is weak (Charman et al., 2010; Charman & Wells, 2012). Similarly, the suggestive influence of intervening lineup photographs (Pezdek & Blandon-Gitlin, 2005) and misinformation (Lane, 2006; Pezdek & Roe, 1995) on eyewitness memory is greater under conditions associated with weaker memory. Such results suggest that feedback might have a stronger effect on subsequent identification attempts when memory for the culprit is weaker. Thus, future research should examine the moderating effect of forensically relevant variables regarding encoding (e.g., brief exposure, divided attention) and testing conditions (e.g., retention interval between witnesses a crime and viewing a lineup). It is also worth considering how feedback might influence subsequent performance in the context of test procedures other than simultaneous lineups, such as sequential lineups and recently-developed identification procedures that differ markedly from typical lineup tests (e.g., Brewer, Weber, & Guerin, 2020).

In addition to exploring hitherto un-researched but realistic scenarios, the mechanisms underpinning the effects of feedback on subsequent decisions must also be elucidated further. Research has so far provided some evidence that feedback influences perceptions of ability and, in turn, performance on memory tasks. However, we do not yet have a comprehensive understanding of this process. For example, consider a comparison between our results and those of Palmer et al. (2010). In our research, disconfirming feedback after an initial pick from a lineup reduced discriminability only when the feedback clearly stated that (a) the person picked earlier was not the culprit and (b) the culprit had been present in the earlier lineup (Experiment 2). When disconfirming feedback might have been interpreted as implying that the culprit was not in the first lineup, it did not influence discriminability (Experiment 1). In contrast, Palmer et al. found that discriminability was reduced by disconfirming feedback that clearly implied the suspect had not been present in the earlier lineup. This comparison suggests that the inferences drawn by witnesses from feedback—and the way these inferences translate to differences in performance—likely differ depending on contextual factors, such as whether a subsequent lineup involves the same culprit or a different culprit. Further insight into such issues may come from probing participants' reactions to feedback, or testing whether the effects of feedback on subsequent identification responses are mediated by differences in perceived memory efficacy. In the current study, we deemed it unwise to include such measures after the first lineup because this might have shaped identification behavior on the second lineup (e.g., Fleet, Brigham, & Bothwell, 1987). Future work might devise a way of examining this mechanism more closely.

References

Beaudoin, M., & Desrichard, O. (2011). Are memory self-efficacy and memory performance related? A meta-analysis. *Psychological Bulletin*, *137*, 211-241.doi:10.1037/a0022106

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

Brehmer, B. (1980). In one word: Not from experience. *Acta Psychologica*, *45*, 223-241. doi:10.1016/0001-6918(80)90034-7

Brewer, N., Weber, N., & Guerin, N. (2020). Police line-ups of the future? *American Psychologist 75*, 76-91. . doi:10.1037/amp0000465

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied*, *12*, 11-30. doi:10.1037/1076-898X.12.1.11

Charman, S. D., Carlucci, M., Vallano, J., & Hyman Gregory, A. R. (2010). The selective cue integration framework: A theory of witness confidence assessment following post-identification feedback. *Journal of Experimental Psychology: Applied*, *16*, 204-218.doi:10.1037/a0019495

Charman, S. D., & Wells, G. L. (2007). Eyewitness lineups: Is the appearance-change instruction a good idea? *Law and Human Behavior*, *31*, 3-22. doi:10.1007/s10979-006-9006-3

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

Crook, T. H., & Larrabee, G. J. (1990). A self-rating scale for evaluating memory in everyday life. *Psychology and Aging*, *5*, 48–57. doi:10.1037/0882-7974.5.1.48

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*, 7-29. doi:10.1177/0956797613504966

Cutler, B. L., Daugherty, B., Babu, S., Hodges, L., & Van Wallendael, L. R. (2009). Creating blind photoarrays using virtual human technology: A feasibility test. *Police Quarterly*, *12*, 289-300. doi:10.1177/1098611109339892

Cutler, B. L., Penrod, S. D., & Stuve, T. E. (1988). Jury decision making in eyewitness identification cases. *Law and Human Behavior*, *12*, 41–56.

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

Duncan, M. J. (2006). *A signal detection model of compound decision tasks* (Technical Report No. TR2006-256). Toronto, ON: Defence Research and Development Canada.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*, 4-58. doi:10.1177/1529100612453266

Fleet, M. L., Brigham, J. C., & Bothwell, R. K. (1987). The confidence-accuracy relationship: The effects of confidence assessment and choosing. *Journal of Applied Social Psychology, 17*, 171-187.

Gallo, D. A., Weiss, J. A., & Schacter, D. L. (2004). Reducing false recognition with criterial recollection tests: Distinctiveness heuristic versus criterion shifts. *Journal of Memory & Language, 51*, 473-493. doi: 10.1016/j.jml.2004.06.002

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.

Gronlund, S. D. (2004). Sequential lineups: Shift in criterion or decision strategy? *Journal of Applied Psychology, 89*, 362-368. doi:10.1037/0021-9010.89.2.362

Hammond, K. R., Summers, D. A., & Deane, D. H. (1973). Negative effects of outcome-feedback in multiple-cue probability learning. *Organizational Behavior and Human Performance, 9*, 30-34. doi:10.1016/0030-5073(73)90034-2

Hertzog, C. (2002). Metacognition in older adults: Implications for application. In T. J. Perfect & B. L. Schwartz (Eds.), *Applied metacognition* (pp. 169–196). New York, NY: Cambridge University Press.

Hertzog, C., & Dixon, R. A. (1994). Metacognitive development in adulthood and old age. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 227–251). Cambridge, MA: MIT Press.

Hertzog, C., Park, D. C., Morrell, R. W., & Martin, M. (2000). Ask and ye shall receive: Behavioral specificity in the accuracy of subjective memory complaints. *Applied Cognitive Psychology, 14,* 257–275. doi:0.1002/(SICI)1099-0720(200005/06)14:3%3C257::AID-ACP651%3E3.0.CO;2-O

Hobson, Z. J., & Wilcock, R. (2011). Eyewitness identification of multiple perpetrators. *International Journal of Police Science and Management*, *13*, 286-296. doi:10.1350/ijps.2011.13.4.253

Horry, R., Brewer, N., Weber, N., & Palmer, M. A. (2015). The effects of allowing a second sequential lineup lap on choosing and probative value. *Psychology, Public Policy, and Law*, *21*, 121-133. doi:10.1037/law0000041

Horry, R., Halford, P., Brewer, N., Milne, R., & Bull, R. (2014). Archival analyses of eyewitness identification test outcomes: What can they tell us about eyewitness memory? *Law and Human Behavior*, *38*, 94-108. doi:10.1037/lhb0000060

Horry, R., Palmer, M. A., & Brewer, N. (2012). Backloading in the sequential lineup prevents within-lineup criterion shifts that undermine eyewitness identification performance. *Journal of Experimental Psychology: Applied*, *18*, 346-360. doi:10.1037/a0029779

Innocence Project (2016). *Eyewitness Misidentification.* Retrieved July 1, from http://www.innocenceproject.org/causes/eyewitness-misidentification/

Kantner, J., & Lindsay, D. S. (2012). Response bias in recognition memory as a cognitive trait. *Memory & Cognition*, *40*, 1163-1177. doi:10.3758/s13421-012-0226-0

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 107-118. doi:10.1037//0278-7393.6.2.107

Kucina, T., Sauer, J. D., Holt, G. A., Brewer, N., & Palmer, M. A. (2020). Refining the blank lineup procedure: How should we instruct eyewitnesses? *Applied Cognitive Psychology*. Advance online publication. doi:10.1002/acp.3719

Lane, S. M. (2006). Dividing attention during a witnessed event increases eyewitness suggestibility. *Applied Cognitive Psychology, 20*, 199–212. doi:10.1002/acp.1177

Lane, S. M., Roussel, C. C., Villa, D., & Morita, S. K. (2007). Features and feedback: Enhancing metamnemonic knowledge at retrieval reduces source-monitoring errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 1131-1142. doi:10.1037/0278-7393.33.6.1131

Lee, J., & Penrod, S. D. (2019). New signal detection theory-based framework for eyewitness performance in lineups. *Law and Human Behavior, 43*, 436-454. doi:10.1037/lhb0000343

Meissner, C. A., Tredoux, C. G., Parker, J. F., & MacLin, O. H. (2005). Eyewitness decisions in simultaneous and sequential lineups: A dual-process signal detection theory analysis. *Memory & Cognition, 33*, 783-792. doi:10.3758/BF03193074

Mickes, L., Flowe, H. D., & Wixted, J. T. (2012). Receiver operating characteristic analysis of eyewitness memory: Comparing the diagnostic accuracy of simultaneous and sequential lineups. *Journal of Experimental Psychology: Applied*, *18*, 361-376. doi:10.1037/a0030609

Mosteller, F., & Tukey, J. W. (1968). Data analysis including statistics. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (pp. 80-203). Reading, MA: Addison-Wesley.

National Research Council (2014). *Identifying the culprit: Assessing eyewitness identification*. Washington, DC: The National Academies Press.

Neil v. Biggers, 409 U.S. 188 (1972).

Palmer, M. A., & Brewer, N. (2012). Sequential lineup presentation promotes less-biased criterion setting but does not improve discriminability. *Law and Human Behavior, 36*, 247-255. doi: 10.1037/h0093923.

Palmer, M. A., Brewer, N., & Weber, N. (2010). Postidentification feedback affects subsequent eyewitness identification performance. *Journal of Experimental Psychology: Applied, 16*, 387-398. doi:10.1037/a0021034

Palmer, M. A., Brewer, N., & Weber, N. (2012). The information gained from witnesses' responses to an initial "blank" lineup. *Law and Human Behavior*, *36*, 439-447. doi:10.1037/h0093939

Palmer, M. A., Brewer, N., Weber, N., & Nagesh, A. (2013). The confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention. *Journal of Experimental Psychology: Applied*, *19*, 55-71. doi:10.1037/a0031602

Perfect, T. J., & Weber, N. (2012). How should witnesses regulate the accuracy of their identification decisions: One step forward, two steps back? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *6*, 1810-1818. doi:10.1037/a0028461.

Pezdek, K., & Blandon-Gitlin, I. (2005). When is an intervening line-up most likely to affect eyewitness identification accuracy? *Legal and Criminological Psychology*, *10*, 247–263. doi:10.1348/135532505X49846

Pezdek, K., & Roe, C. (1995). The effect of memory trace strength on suggestibility. *Journal of Experimental Child Psychology*, *60*, 116–128. doi:10.1006/jecp.1995.1034

Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. T*rends in Cognitive Sciences*, *15*, 20–27. doi:10.1016/j.tics.2010.09.003

Sauer, J. D., Brewer, N., Zweck, T., & Weber, N. (2010). The effect of retention interval on the confidence-accuracy relationship for eyewitness identification. *Law and Human Behavior*, *34*, 337–347.doi:10.1007/s10979-009-9192-x

Sauer, J. D., Palmer, M. A., & Brewer, N. (2019). Pitfalls in using eyewitness confidence to diagnose the accuracy of an individual identification decision. *Psychology, Public Policy, and Law*, *25*, 147-165. doi:10.1037/law0000203

Schacter, D. L., Israel, L., & Racine, C. A. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language*, *40*, 1-24. doi:10.1006/jmla.1998.2611

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

Smalarz, L., Kornell, N., Vaughn, K. E., & Palmer, M. A. (2019). Identification performance from multiple lineups: Should eyewitnesses who pick fillers be burned? *Journal of Applied Research in Memory and Cognition*, *8*, 221-232. doi:10.1016/j.jarmac.2019.03.001

Smalarz, L., & Wells, G. L. (2014). Confirming feedback following a mistaken identification impairs memory for the culprit. *Law and Human Behavior*, *38*, 283-292. doi:10.1037/lhb0000078

Smith, A. M., Wells, G. L., Smalarz, L., & Lampinen, J. M. (2018). Increasing the similarity of lineup fillers to the suspect improves the applied value of lineups without improving memory performance: Commentary on Colloff, Wade, & Strange (2016). *Psychological Science*, *29*, 1548-1551. doi:10.1177/0956797617698528

Smith, A. M., Yang, Y., & Wells, G. L. (2020). Distinguishing between investigator discriminability and eyewitness discriminability: A method for creating full Receiver Operating Characteristic curves of lineup identification procedures. *Perspectives on Psychological Science*, *15*, 589 – 607. doi:10.1177/174569162-9-2426

Sporer, S. L., Penrod, S. D., Read, J. 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

Steblay, N. K., Tix, R. W., & Benson, S. L. (2013). Double exposure: The effects of repeated identification lineups on eyewitness accuracy. *Applied Cognitive Psychology*, *27*, 644-654. doi:10.1002/acp.2944

Steblay, N. K., Wells, G. L., & Douglass, A. B. (2014). The eyewitness post identification feedback effect 15 years later: Theoretical and policy implications. *Psychology, Public Policy, and Law*, *20*, 1-18. doi:10.1037/law0000001

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1379–1396. doi:10.1037/0278-7393.24.6.1379

Tryon, W. W. (2001). Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: An integrated alternative method of conducting null hypothesis statistical tests. *Psychological Methods*, *6*, 371-386.doi:10.1037/1082-989X.6.4.371

Vitriol, J. A., Appleby, J., & Borgida, E. (2018). Racial bias increases false identification of black suspects in simultaneous lineups. *Social Psychological and Personality Science*, *10*, 722-734. doi:10.1177/1948550618784889

Weber, N., & Brewer, N. (2006). Positive versus negative face recognition decisions: Confidence, accuracy, and response latency. *Applied Cognitive Psychology, 20*, 17-31.doi:10.1002/acp.1166

Weber, N. and Perfect, T. J. (2012). Improving eyewitness identification accuracy by screening out those who say they don't know. *Law and Human Behavior*, *36,* 28-36. doi:10.1037/h0093976

Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology*, *14*, 89-103. doi:10.1111/j.1559-1816.1984.tb02223.x

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 postidentification 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., Kovera, M. B., Douglass, A. B., Brewer, N., Meissner, C., & Wixted, J. (In press, accepted November 20, 2019). Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. *Law and Human Behavior*.

Wells, G. L., Small, M., Penrod, S., 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.

Wixted, J. T., Vul, E., Mickes, L. & Wilson, B. W. (2018). Models of lineup memory. *Cognitive Psychology, 105*, 81-114. doi:10.1016/j.cogpsych.2018.06.001

Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest, 18*, 10-65. doi:10.1177/1529100616686966

Table 1

Percentages and Frequencies of Identification Responses for Target-Present and Target-Absent Lineups (Second Lineup Only) by Condition in Experiment 1

	Target-present lineups						
Initial decision & feedback condition	Correct identification		Filler identification		Lineup rejection		Total
	%	*n*	%	*n*	%	*n*	*n*
Single lineup	33.3	12	19.4	7	47.2	17	36
Initial choosers							
No feedback	40.6	13	18.8	6	40.6	13	32
Feedback	30.4	7	13.0	3	56.5	13	23
Initial non-choosers							
No feedback	45.0	18	15.0	6	40.0	16	40
Feedback	27.3	9	33.3	11	39.4	13	33
Overall	36.0	59	20.1	33	43.9	72	164

	Target-Absent Lineups				
	Filler Identification		Correct Rejection		Total
	%	*n*	%	*n*	*n*
Single lineup	19.4	7	80.6	29	36
Initial choosers					
No feedback	46.2	12	53.8	14	26
Feedback	23.5	4	76.5	13	17

Initial non-choosers					
No feedback	32.6	15	67.4	31	46
Feedback	48.8	20	51.2	21	41
Overall	34.9	58	65.1	108	166

Table 2

Estimates of d′ and c for each condition in Experiments 1 and 2, calculated with an SDT-CD integration model. Non-overlapping inferential confidence intervals (95% ICIs) indicate a difference between conditions at the alpha = .05 level.

Group and condition	*d′*	SE	95% ICI	*c*	SE	95% ICI
Experiment 1						
Single lineup	1.68	0.33	[1.21, 2.15]	-0.12	0.18	[-0.47, 0.24]
Initial non-choosers						
No feedback	1.92	0.30	[1.49, 2.35]	-0.47	0.17	[-0.71, -0.23]
Feedback	0.97	0.32	[0.52, 1.42]	-0.41	0.18	[-0.67, -0.15]
Initial choosers						
No feedback	1.59	0.33	[1.11, 2.06]	-0.53	0.20	[-0.81, -0.24]
Feedback	1.71	0.48	[1.02, 2.40]	-0.05	0.25	[-0.40, 0.31]
Experiment 2						
Initial choosers						
No feedback	1.56	0.21	[1.26, 1.86]	-1.14	0.15	[-1.35, -0.93]
Feedback	0.90	0.23	[0.57, 1.22]	-0.86	0.15	[-1.07, -0.64]

Note. The 95% ICIs for Experiment 1 were calculated using an average *E* parameter, enabling comparisons between any pair of *d′* values or *c* values within that experiment.

Table 3

Percentages and Frequencies of Identification Responses for Target-Present and Target-Absent Lineups (Second Lineup Only) by Condition in Experiment 2

	Target-present lineups						
	Correct identification		Filler identification		Lineup rejection		Total
Feedback condition	%	*n*	%	*n*	%	*n*	*n*
Initial choosers							
No feedback	52.6	30	28.1	16	19.3	11	57
Feedback	32.7	18	38.2	21	29.1	16	55
Overall	42.9	48	33.0	37	24.1	27	112

	Target-Absent Lineups				
	Filler Identification		Correct Rejection		Total
	%	*n*	%	*n*	*n*
Initial choosers					
No feedback	67.9	38	32.1	18	56
Feedback	72.9	43	27.1	16	59
Overall	70.4	81	29.6	34	115

Figure Caption

Figure 1. Mean confidence for correct and incorrect identification decisions (with Cohen's *d* effect sizes) for each condition in Experiments 1 and 2. Error bars indicate 95% confidence intervals.

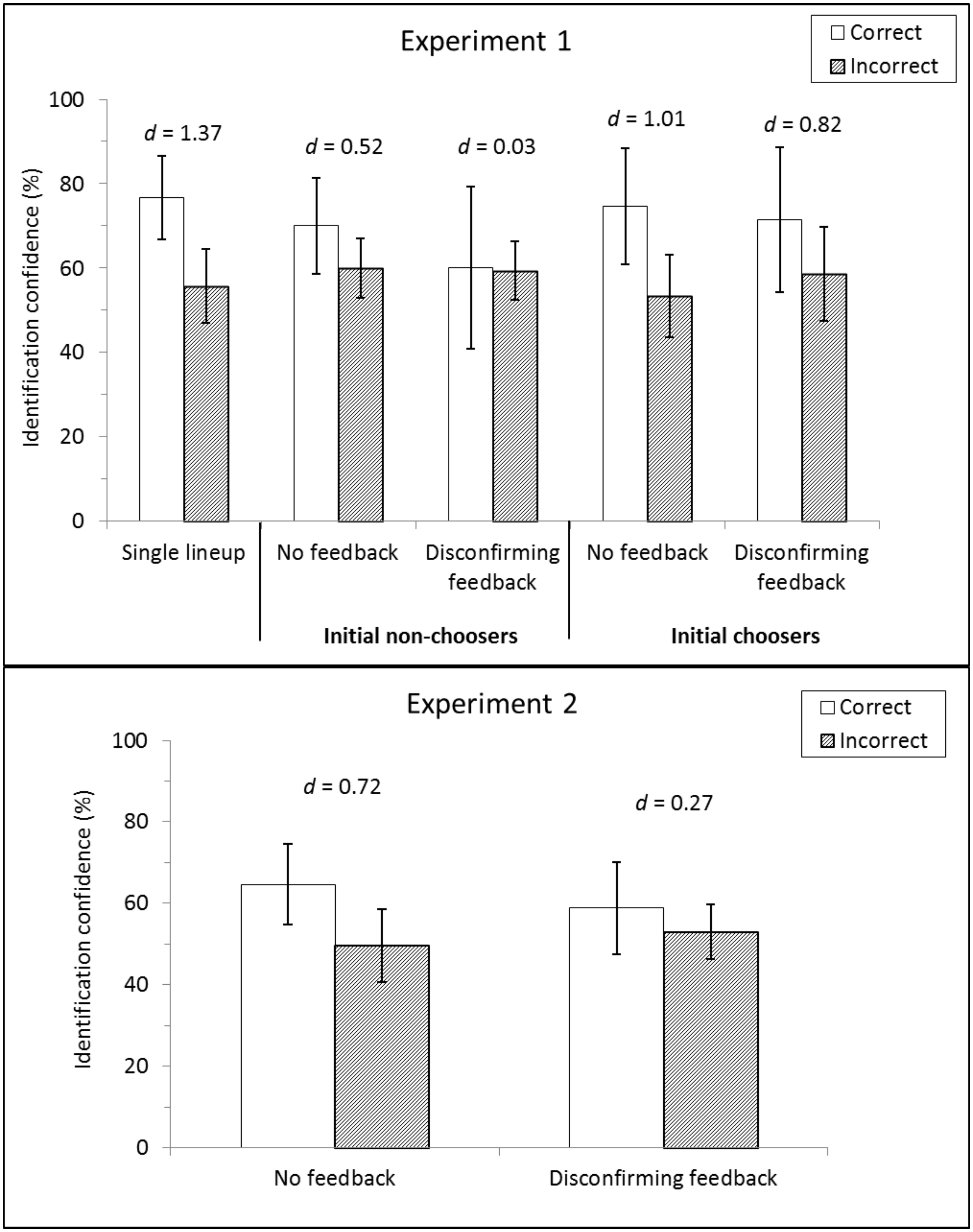
Experiment 1
Correct
Incorrect
Identification confidence (%)
100
80
60
40
20
0
d = 1.37
d = 0.52
d = 0.03
d = 1.01
d = 0.82
Single lineup
No feedback
Disconfirming feedback
No feedback
Disconfirming feedback
Initial non-choosers
Initial choosers
Experiment 2
Correct
Incorrect
Identification confidence (%)
100
80
60
40
20
0
d = 0.72
d = 0.27
No feedback
Disconfirming feedback

Figure 1.

Supplemental Material

Table S1: SDT-CD Independent Observations Model

Estimates of d' and c for each condition in Experiments 1 and 2, calculated with an SDT-CD independent-observations model. Non-overlapping inferential confidence intervals (95% ICIs) indicate a difference between conditions at the alpha = .05 level.

Group and condition	*d'*	SE	95% ICI	*c*	SE	95% ICI
Experiment 1						
Single lineup	1.60	0.26	[1.23,1.97]	0.99	0.13	[0.81,1.17]
Initial non-choosers						
No feedback	1.69	0.23	[1.36, 2.02]	0.76	0.12	[0.59, 0.92]
Feedback	0.94	0.28	[0.54, 1.34]	0.81	0.14	[0.61, 1.00]
Initial choosers						
No feedback	1.42	0.27	[1.03,1.80]	0.72	0.14	[0.52, 0.92]
Feedback	1.50	0.37	[0.98, 2.03]	1.07	0.18	[0.82, 1.32]
Experiment 2						
Initial choosers						
No feedback	1.46	0.19	[1.20,1.72]	0.30	0.11	[0.15.0.45]
Feedback	0.85	0.21	[0.57,1.14]	0.55	0.11	[0.39,0.70]

Note. The 95% ICIs for Experiment 1 were calculated using an average *E* parameter, enabling comparisons between any pair of *d′* values or *c* values within that experiment.

Model fitting

Integration model. Model fit statistics, and model-generated and observed response proportions appear in Table S2. For Experiment 1, the best-fitting model in each condition represented a good fit to the observed data (all G_{total} values < 1.06; all p values > .78). For Experiment 2, fit between the observed and model-generated response probabilities was good for the no-feedback condition ($G < 1$, $p = .83$). The model did not fit the data quite so well for the feedback condition, although the likelihood test result associated with the latter comparison did not approach statistical significance ($G = 3.32$, $p = .34$).

Independent observation model. Model fit statistics, and model-generated and observed response proportions appear in Table S3. The absence of any statistically significant discrepancies between model-generated and observed response proportions indicates that the independent observations model provided an adequate fit to the data. However, a comparison of *G* values between Tables S2 and S3 indicates that the integration model fit the data better than the independent observations model.

Table S2: Model fit statistics for SDT-CD integration model

Observed and Model-Generated Response Probabilities, Model Fit Statistic (G_{total}), and Associated Statistical Significance Test Value (p) for Each Condition in Experiments 1 and 2

Group and	Observed			Model				
condition	CID	FID	FA	CID	FID	FA	G_{total}	*p*
Experiment 1								
Single lineup	.33	.19	.19	.35	.17	.23	0.72	.87
Initial non-choosers								
No feedback	.45	.15	.33	.44	.16	.31	0.10	.99

Feedback	.27	.33	.49	.26	.34	.47	0.13	.99
Initial choosers								
No feedback	.41	.19	.46	.38	.21	.39	1.05	.79
Feedback	.30	.13	.24	.29	.14	.21	0.13	.99
Experiment 2								
Initial choosers								
No feedback	.53	.28	.68	.51	.30	.64	0.90	.83
Feedback	.33	.38	.73	.29	.42	.66	3.32	.34

Notes. CID = correct identification from target-present lineup; FID = filler identification from target-present lineup; FA = positive identification from target-absent lineup (false alarm).

Table S3: Model fit statistics for SDT-CD independent observations model

Observed and Model-Generated Response Probabilities, Model Fit Statistic (G_{total}*), and Associated Statistical Significance Test Value (p) for Each Condition in Experiments 1 and 2*

Group and condition	Observed			Model				
	CID	FID	FA	CID	FID	FA	G_{total}	*p*
Experiment 1								
Single lineup	.33	.19	.19	.34	.19	.20	0.02	.99
Initial non-choosers								
No feedback	.45	.15	.33	.40	.20	.29	1.46	.69
Feedback	.27	.33	.49	.26	.35	.47	0.17	.98
Initial choosers								

No feedback	.41	.19	.46	.35	.25	.38	2.15	.54
Feedback	.30	.13	.24	.27	.17	.19	0.71	.87
Experiment 2								
Initial choosers								
No feedback	.53	.28	.68	.48	.32	.63	2.12	.55
Feedback	.33	.38	.73	.28	.43	.66	3.36	.34

Notes. CID = correct identification from target-present lineup; FID = filler identification from target-present lineup; FA = positive identification from target-absent lineup (false alarm).