Psychological Science (in press)

Identifying the Bad Guy in a Lineup using Deadlined Confidence Judgments

Neil Brewer, Nathan Weber, and David Wootton

School of Psychology

Flinders University

D. Stephen Lindsay

University of Victoria

Neil Brewer, Nathan Weber, and David Wootton, School of Psychology, Flinders University, Adelaide, Australia. D. Stephen Lindsay, Department of Psychology, University of Victoria, Canada.

This research was supported by Australian Research Council Grant DP1093210 and a Flinders Research Grant. Thanks to Nicole Reid and Rachel Hiller who assisted with data collection.

Correspondence concerning this article should be addressed to Neil Brewer, School of Psychology, Flinders University, GPO Box 2100, Adelaide, South Australia, Australia.

E-mail: neil.brewer@flinders.edu.au

Summary

Eyewitness identification tests often culminate in witnesses not picking the culprit or identifying innocent suspects. We tested a radical alternative to the traditional lineup. Rather than identifying someone, witnesses made confidence judgments under a short deadline about each member (shown sequentially) being the culprit. Two experiments compared this procedure with a conventional sequential lineup, using a five-minute retention interval. A third experiment used a retention interval of one week. A classification algorithm identified confidence criteria that optimally discriminated accurate from inaccurate decisions, with decision accuracy 21-66% higher under the deadline procedure. Confidence profiles across lineup stimuli were more informative than identification decisions about the likelihood that an individual witness recognized the culprit or correctly rejected the lineup. Large differences between the maximum and next highest confidence value signaled very high accuracy. Future support for this procedure across varied conditions would highlight a viable alternative to the problematic traditional lineup procedures.

Eyewitness identification evidence often proves to be very persuasive in the courtroom. Indeed, eyewitness testimony is sometimes the only evidence available, with convictions determined by such evidence (Semmler, Brewer, & Douglass, 2011; Wells et al., 1998). Yet, the fallibility of eyewitness identification evidence – manifested in two striking response patterns – has been consistently highlighted in laboratory and field studies. First, witnesses sometimes identify an innocent suspect as the culprit, paving the way for an erroneous conviction. The potency of mistaken identifications is dramatically highlighted by DNA exoneration cases: The US Innocence Project website (2012) reveals that mistaken identifications were important in the wrongful conviction of more than 70% of the 289 DNA exoneration cases documented to date. Second, witness fallibility can manifest in the rejection of a police lineup where the suspect actually is the culprit, possibly leading to the culprit going free. Field and laboratory studies indicate that (incorrect) lineup rejections are commonplace (Pike, Brace, & Kynan, 2002; Steblay, Dysart, Fulero, & Lindsay, 2001). While various procedural manipulations have been shown to reduce the likelihood of false identifications (Brewer & Palmer, 2010; Wells, Memon, & Penrod, 2006), the frequency with which these two problematic responses occur highlights the need for an alternative to the traditional eyewitness identification test that has survived largely intact since the 19th century. As Wells, Memon, and Penrod (2006) noted:

> "…. what if the lineup had never existed and the legal system turned to psychology to determine how information could be extracted from eyewitnesses' memories? … Operating from scratch, it seems likely that modern psychology would have developed radically different ideas …. brain-activity measures, eye movements, rapid displays of faces, reaction times, and other methods for studying memory might have been developed instead of the traditional lineup. …. it is possible to

imagine a future science of eyewitness evidence that is radically different from the methods used today (pp. 68-69)."

Here we report three experiments demonstrating the effectiveness of a radical alternative to the traditional lineup.

A weakness of the traditional identification test or lineup lies in the fact that it requires a witness to make a single decision about a lineup. The witness examines the lineup (comprising the suspect and a number of foils who match the suspect's description and/or look similar to the suspect) and either makes a choice from the lineup (i.e., picks the suspect or a known-innocent foil) or rejects the lineup (i.e., says the culprit is not there or can't decide). Perhaps if every witness was guaranteed to have a high quality memorial representation of the offender, these might be reasonable expectations. But factors such as a limited exposure to the culprit and a delay to the identification test (Brewer & Wells, 2011) often mitigate against such a memory, so the witness has to weigh up whether a positive or negative decision is more appropriate based on the limited memorial evidence available. The witness's decision criterion is also vulnerable to influence from an array of social cues (Brewer & Palmer, 2010; Wells et al., 2006) that bias them towards a positive identification. Or, metacognitive cues (e.g., "I saw him for a long time") may persuade the witness that (s)he should have a strong memory and to set a strict decision criterion demanding strong evidence for selecting a stimulus because the target stimulus should be familiar (Morell, Gaitan, & Wixted, 2002). Finally, the perceived significance of the identification decision may encourage the activation of heuristics that bias criterion setting. Consider, for example, a victim of an assault who is viewing a lineup and has plenty of time to reflect on the decision. Weighing heavily are likely to be considerations such as "If I pick this person he'll probably get a long jail sentence, so I'd better be right" or "If I don't pick this person, a dangerous man may well go free, so I'd better make sure before I say he's not the one." Such thoughts are also likely to promote reflection on factors such as the quality of the viewing conditions, attentional constraints, and so on. In sum, various social, metacognitive and heuristic influences can produce overly lax or unduly stringent response criteria (Clark, 2005), contributing to false identifications or rejections.

Using a variety of encoding and test stimuli we tested a novel procedure for accessing eyewitness memory while minimizing the influence of strategic processes that often affect witnesses' decision criteria in a traditional identification test. There are two key components to our procedure. The first involves having the witness rate the degree of match between the culprit and each lineup member (presented sequentially). The second involves requiring the witness to perform this matching process under severe constraints on processing time. Rather than requiring the witness to make a Yes/No identification decision, the first component of our procedure requires confidence judgments about the likelihood that each of an array of faces (presented serially) – including a target face (i.e., the culprit or an innocent suspect) and a number of foils – is that of the person who committed the crime. Thus, we use confidence judgments as an index of memory strength. Although retrospective confidence judgments are malleable and, hence, can be poor indicators of eyewitness identification accuracy (Brewer, 2006; Douglass & Steblay, 2006), this problem can be ameliorated by collecting confidence measures in a non-biased way prior to any post-identification social interaction with other witnesses or lineup administrators and under conditions that limit opportunities for deliberation (Brewer & Weber, 2008; Brewer & Wells, 2006). This approach is consistent with theory and research in recognition memory which indicates that confidence judgments can index memory strength, discriminating previously seen from unseen stimuli (Bernbach, 1971; Cleary & Greene, 2000; Wickelgren & Norman, 1966), although these theories were designed to account for decision making contexts much less complex than the present one. Similarly, as Sauer, Brewer, and Weber (2008) argued, this approach aligns with metamemory research indicating that people have at least partial access to retrieval information that supports memory-based decisions even in the absence of a positive decision (cf. Koriat, 1993). However, a significant departure in our present research is that, rather than witnesses classifying each stimulus as old/new and assigning a confidence value, they simply provide a confidence rating for each lineup stimulus.

Requiring witnesses to make a series of confidence assessments about individual stimuli – without any accompanying Yes/No identification decision – should negate the influences of some of the biases that disturb criterion setting. Moreover, even if some of these social and metacognitive influences modulate confidence assessments, they should exert a uniform influence across the set of stimuli, thereby still allowing for confidence differences that track variations in the match between each probe stimulus and the witness's memory. In previous research (Sauer et al., 2008) we showed a confidence criterion could be identified at the group level that discriminated whether an offender's face had or had not been recognized better than the traditional identification decisions.

Here, to minimize the contribution of any strategic influences on such confidence judgments, ratings were made under a response-signal deadline (Brewer & Smith, 1990; Wickelgren, 1977). Specifically, witnesses made the confidence judgment about each face within 3 seconds of it appearing. Strong memory traces are accessed more rapidly than weaker traces (Murdock, 1985), and accurate eyewitness identifications are significantly faster than inaccurate ones (Brewer, Caon, Todd, & Weber, 2006; Weber, Brewer, Wells, Semmler, & Keast, 2004). Thus, confidence judgments obtained under constrained processing time should further minimize the contribution of those factors that can disturb criterion setting for identification decisions. We speculated that the response deadline would have little effect on rapid-memory based responses to the target but would truncate the more time-consuming processes that contribute to false identifications. Thus, a strong memory for the target should still result in a high level of confidence, whereas confidence for unfamiliar faces should be low. Consistent with this argument are the findings in the recognition memory literature that familiarity judgments, including those for face stimuli (i.e., judgments that make a major contribution to confidence estimates, Wixted, 2007) – are usually made extremely rapidly (Bruce & Young, 1998; Johnson, Kounios, & Reeder, 1994; Yonelinas & Jacoby, 1994). Moreover, short deadlines minimize the likelihood that metacognitive cues or heuristics will distort the confidence values reported (especially in an item-specific manner). Thus, at short deadlines the discrepancy between confidence values for previously seen versus unfamiliar faces should be large when the witness's memory is strong (and hence likely accurate). Put another way, large confidence discrepancies between one line-up face and all the others would suggest a strong memory match and a high likelihood that the face attracting the standout, high-confidence value is the culprit; small discrepancies would suggest the opposite.

Three experiments (each using multiple stimuli) compared the deadline confidence procedure with the traditional sequential lineup format (i.e., control condition) in which witnesses viewed one face at a time without knowing how many faces would be shown (cf. Lindsay & Wells, 1985; Steblay, Dysart, & Wells, 2011). In the first two experiments, the retention interval between encoding and test was five minutes; in the third experiment it was one week. To compare the effectiveness of the two procedures, first a classification algorithm identified confidence criteria that optimally discriminated accurate and inaccurate decisions, providing the crucial comparison of overall classification accuracy produced by the two lineup conditions. While this provides an index of the efficacy of the procedures at the group level, it does not inform the key forensic judgment about the likelihood that an individual witness is accurate. Consequently, the second approach used a profile analysis that indicates the likelihood that any individual's pattern of confidence judgments reflects an accurate judgment about the target face.

Method

Participants and Design

Experiment 1 included 494 undergraduate and community participants (219 male, 275 female) aged 16-54 years (*M* = 20.1, *SD* = 6.7). Experiment 2 included 294 participants (117 male, 177 female) aged 16-53 years (*M* = 22.1, *SD* = 6.4). Experiment 3 included 117 participants (48 male, 69 female) aged 16-60 years (*M* = 26.3, *SD* = 9.0)

Participants were randomly assigned to either the deadline confidence or standard sequential lineup condition, with lineup condition and culprit presence/absence varied evenly between-and within-subjects, respectively.

Materials

PCs presented the stimulus events, distracter movie and lineups; recorded responses; and controlled the deadline. In Experiments 1 and 2, filler movies of 5-minute TV programs were used during the retention interval between viewing the stimulus event and the lineup. In Experiment 3 the participant returned to the laboratory as near as possible to one week after viewing the stimulus event (range = 3-13 days, *M* = 6.7, *SD* = 1.2).

In Experiment 1 the two movies depicted two different simulated non-violent crimes. In Experiments 2 and 3 the movies depicted four different stimulus events (involving four different people); after they had viewed each movie, witnesses were advised that the target was a suspect in a crime that had occurred in the vicinity. The movies were selected from an array used in our laboratory because our archival data show that they produce varied patterns of identification test choosing and accuracy, thus allowing a test of the procedures across multiple stimuli of varying difficulty, just as happens in the real world. Presentation order was counterbalanced.

The 12 lineup photos each had an on-screen size of 8cm × 6cm, and provided a frontal view from the chest up. The foils (and the target's replacement for each target-absent lineup) were selected using a match-description strategy; all foils matched a modal free-recall description of the target provided by independent observers of each stimulus event (Wells, 1993). Thus, if the target was described as male, 20-30 years of age, short darkish hair, pale complexion, all lineup members matched that description.

Procedure

Participants sat alone at a computer in a cubicle and received the instruction: "You are going to be shown a short film. Pay close attention to it because you will be asked some questions afterwards." Participants were then presented with: "When you are ready to watch the film click the "Next" button. A count-down appeared in the centre of the screen and, after a short delay, the film started. After participants watched the first movie, another screen appeared: "Before completing the questions about the film you have just seen, we would like you to watch another film. Once again, please pay close attention." After watching two movies (Experiment 1) or four movies (Experiment 2), participants watched a distracter film for 5 minutes. Experiment 3 participants watched four movies and returned to the laboratory one week later.

For the lineup phase, the screen displayed instructions for viewing the lineup and making responses. In both lineup conditions participants viewed the 12 lineup photos in serial order and were not informed about how many photos were in the lineup. The first 4 faces were the same for all participants in both conditions and, though matching the culprit's description, were deliberately chosen because they were not very plausible selections. Their purpose was to acclimatize the witness to the deadline procedure (and to duplicate that in the control condition), although witnesses were not informed that these were only warm-up trials. The order of the next 8 faces (the actual trials) was random.

In the deadline condition each face was on the screen for 3 seconds. After 2 seconds a buzzer indicated that the confidence judgment had to be completed within the next second. This involved clicking one of 11 on-screen buttons spanning *100% confident that this is the culprit* through to *0% absolutely certain this is not the culprit*. If the confidence judgment was not made before the deadline (around 2.5% of trials), the face disappeared and the next face appeared. Note that pilot work with shorter deadlines produced many failures to respond within the deadline.

The control condition involved a standard sequential lineup of 12 faces and participants had as much time as they needed to view each photo. A Yes/No decision was required for each face. If the witness chose a particular face, the lineup ended. If the witness rejected a face, the next face appeared, and so on, until a choice was made or the lineup ended. (In Lindsay and Wells', 1985, original sequential lineup study the lineup continued to the end even if the witness made a selection; in many other studies the lineup ended when witnesses made a selection. Allowing witnesses only one pass through a sequential lineup has been a deliberate strategy designed to try to 'force' witnesses to make absolute rather than relative judgments and hence overcome the tendency for witnesses to make a choice even when no strong match to memory is found.) When the lineup ended a new screen asked participants to express how confident they were in their final identification decision using 1 of 11 on-screen buttons (0%-100%). The same screen was presented to those who rejected all lineup members.

Results

Two approaches were used to explore the efficacy of the deadline confidence procedure: (1) A comparison of overall (i.e., group-level) classification accuracy based on optimally discriminating confidence criteria, and (2) A profile analysis on confidence judgments across stimuli at the individual witness level to highlight patterns associated with high versus low accuracy.

The Comparison of Deadline and Control Condition Accuracy

We used Sauer et al.'s (2008) hierarchical algorithm to infer a decision from the confidence ratings for the deadline condition (i.e., in which the witness didn't make a Yes/No decision) and thereby obtain a measure of accuracy. First, when a participant provided a single maximum (Max) confidence value – one value for a lineup member that was higher than all of the others – this was treated initially as a positive decision (i.e., a possible selection of that lineup member). The absence of a single Max value was treated as a negative decision (or no selection) and, therefore, indicative of the suspect's innocence. Cases with a Max value were then separated according to whether the Max value referred to the suspect or a foil. Suspect and foil Max values were then compared to one of two separate criteria derived from all participants' data. Both the suspect and foil criteria were calculated as the Max value (i.e., maximum confidence value) that optimally discriminated accurate from inaccurate decisions across participants. For example, if a suspect Max confidence value of 70% produced the highest proportion of correct decisions across all participants, then 70% became the suspect confidence criterion. The confidence criterion algorithm maximizes the number of accurate decisions without any attention to classification as hits or correct rejections.

Suspect Max values equaling or exceeding the suspect confidence criterion were classed as positive identifications of the suspect. Foil Max values equaling or exceeding the foil confidence criterion were counted as lineup rejections because, in the forensic context, using a (recommended) single-suspect lineup means that lineup foils are known (in advance) to be innocent. Thus, identification of a foil as the best match provides evidence that the suspect is innocent (Clark & Wells, 2008; Wells & Olson, 2002). Suspect or foil Max values falling below their confidence criterion were classed as inconclusive responses. This approach resulted in one of three classifications: (1) The suspect was identified as the culprit, (2) The suspect was innocent, or (3) Inconclusive (such classifications are included in the denominator when calculating classification accuracy).

Table 1 shows the decision accuracy rates (and inferential statistics) for the control condition and those produced by the application of the hierarchical classification algorithm to the deadline confidence condition for each stimulus event/lineup in Experiments 1, 2 and 3. Overall decision accuracy was significantly higher (21%, 31%, 66%) under the deadline procedure in all three experiments. The Cohen's *w* measures indicate meaningful effect sizes for 7 of the 10 stimulus sets and for the overall contrasts for each experiment. (Proportions correct for identification and rejection response classifications are reported in the Supplementary Materials.)

Profiles of Individual Confidence Judgments Indicative of Accuracy

Knowing that, at the group level, this procedure produces higher decision accuracy than a conventional lineup across a wide array of encoding and test stimuli is extremely valuable from the perspective of evaluating the efficacy of the procedure. However, in an actual criminal investigation the focus is on the likely accuracy of an individual witness's decision. Accordingly, we conducted a profile analysis that highlights when an individual set or pattern of confidence judgments is likely/not likely to indicate an accurate discrimination of a previously seen face. We examined accuracy as a function of *Max*–

next highest confidence value. (Various other permutations, such as *Max – average of rest*, produced no better outcomes.) This indicates accuracy classification rates when the Max confidence value was 100% and the next highest value was 0%; when the Max was 100% and the next highest value was 10%, 20%, 30%, etc; when the Max was 90% and the next highest value was 0%, 10%, 20%, etc; and so on. Thus, using this procedure it is possible to compare classification accuracy, given particular profiles of confidence judgments, with decision accuracy for the control condition. Note that for target-present lineups, Max values assigned to foils were excluded as, in appropriately conducted single-suspect lineups, foils are known in advance to be innocent. Also, unlike the hierarchical algorithm described above, which requires that an innocent suspect be designated in a target-absent lineup, the profile analysis classifies the presence of any single Max value given for a target-absent lineup as an inaccurate decision, no matter how low the Max is (i.e., it applies a conservative test).

As shown in Table 2, when the *Max–next highest* confidence discrepancy was large, the profile analysis produced accuracy classification rates that were markedly higher than the control condition decision accuracy (0.51, 0.59 and 0.38 for Experiments 1, 2 and 3, respectively). Accuracy was extremely high when the discrepancies were near the top of the range. Only when the discrepancy fell to around 30-50% did accuracy approach the levels of the control procedure. Moreover, when the discrepancy was very small, accuracy was very low.

Discussion

In two experiments including six different sets of encoding and test stimuli, the deadline confidence procedure produced significantly higher classification accuracy than a traditional sequential lineup control condition. In a third experiment this finding was replicated when the retention interval was extended to one week. Moreover, the profile analysis identified patterns of confidence judgments across lineup stimuli that were more informative than a witness's lineup decision about the likelihood that the witness has recognized the culprit or that the culprit is not in the lineup. Large differences between the maximum and next highest confidence rating (i.e., 70-100%) denoted very high accuracy; in contrast, small differences (≤ 20%) signaled low accuracy. In other words, an individual's confidence profile offers probative information that is not provided by the witness's decision on the standard identification test. As noted earlier, these results are compatible with extrapolations from theory underpinning basic research on recognition memory and metamemory, demonstrating the heuristic value of such theoretical perspectives for investigating decision making in everyday contexts of far greater complexity than those the theories were originally developed to accommodate. Ultimately, a formal model relating accuracy, confidence and response time in this context would be desirable, with existing recognition memory models (e.g., Ratcliff & Starns, 2009; Van Zandt, 2000) providing potential starting points for the development of such a model.

There remain, of course, some obvious significant challenges for future research using this procedure. One is to obtain converging evidence for its efficacy under a variety of forensically relevant conditions (e.g., varied encoding conditions, variations in lineup discriminability). Another is to demonstrate the efficacy of the procedure under conditions where familiarity-based confidence judgments might be considered problematic. For example, if – prior to viewing the lineup – a witness had been shown mugshots that included one of the lineup foils, it might be expected that familiarity for that foil would be higher and confidence would be inflated. This could undermine the effectiveness of the confidence procedure when witnesses were exposed to a culprit-absent lineup containing the foil. It could also undermine the procedure by reducing the

Max–next highest confidence discrepancy when witnesses were exposed to a lineup containing the target and previously seen foil. That is, it should provide a tough test of the deadline procedure. Nevertheless, we suggest that should a single memory strength judgment – incorporating familiarity and contextual information – underpin the deadline confidence judgment (e.g., as suggested by Wixted, 2007), the efficacy of the procedure should be ensured. In contrast, if witnesses had opportunity for reflection (i.e., in the absence of a deadline), previous exposure to one of the lineup foils might produce a sense of familiarity which, upon reflection, could be attributed to the face being that of the culprit. This last issue suggests another important question to be resolved: namely, is the deadline critical for curtailing metacognitive activity that distorts confidence assessments or is it unimportant?

Clearly, adoption of a procedure similar to this deadline confidence procedure would be perceived as a radical change, and one that would not be easily sold to the criminal justice community (cf. Brewer & Wells, 2011). Thus, further research confirming its efficacy and resolving the optimal deadline and confidence scale formats under a variety of conditions such as those outlined above will be crucial for demonstrating that finally there may exist a superior procedure to the age-old and problematic identification test.

References

Bernbach, H. A. (1971). Strength theory and confidence ratings in recall. *Psychological Review, 78*, 338-340.

Brewer, N. (2006). Uses and abuses of eyewitness identification confidence. *Legal and Criminological Psychology*, *11*, 3-24.

Brewer, N., Caon, A., Todd, C., & Weber, N. (2006). Eyewitness identification accuracy and response latency. *Law and Human Behavior, 30*, 31-50.

Brewer, N., & Palmer. M. A. (2010). Eyewitness identification tests. *Legal and Criminological Psychology, 15*, 77-96.

Brewer, N., & Smith, G. A. (1990). Processing speed and mental retardation: Deadline procedures indicate fixed and adjustable limitations. *Memory and Cognition, 18,* 443-450.

Brewer, N., & Weber, N. (2008). Eyewitness confidence and latency: Indices of memory processes not just markers of accuracy. *Applied Cognitive Psychology, 22*, 827-840.

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.

Brewer, N., & Wells, G. L. (2011). Eyewitness identification. *Current Directions in Psychological Science, 20*, 24-27.

Bruce, V., & Young, A. (1998). *In the eye of the beholder: The science of face perception*. New York: Oxford University Press.

Clark, S. E. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law and Human Behavior, 29*, 395-424.

Clark, S. E., & Wells, G. L. (2008). On the diagnosticity of multiple-witness identifications. *Law and Human Behavior, 32*, 406-422.

Cleary, A. M., & Greene, R. L. (2000). Recognition without identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1063-1069.

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

Innocence Project. *Innocence Project*. Retrieved February 11, 2012, from http://www.innocenceproject.org (2012).

Johnson, M. K., Kounios, J., & Reeder, J. A. (1994). Time-course studies of reality monitoring and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1409-1419.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100*, 609-639.

Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identification from lineups: Simultaneous versus sequential lineup presentations. *Journal of Applied Psychology, 70*, 556-564.

Morrell, H. E. R., Gaitan, S., & Wixted, J. T. (2002). On the nature of the decision axis in Signal-Detection-Based models of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 1095-1110.

Murdock, B. B. (1985). An analysis of the strength-latency relationship. *Memory and Cognition, 13*, 511-521.

Pike, G., Brace, N., & Kynan, S. *The visual identification of suspects: Procedures and practice* (Briefing Note 2/02). London: Home Office. (2002).

Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological Review, 116,* 59-83.

Semmler, C., Brewer, N., & Douglass, A. B. (2011). Jurors believe eyewitnesses. In B. L. Cutler (Ed.), *Conviction of the innocent: Lessons from psychological research* (pp. 185-209). Washington, DC: APA Books.

Van Zandt, T. (2000). ROC curves and confidence judgments in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 582-600.

Sauer, J. D., Brewer, N., & Weber, N. (2008). Multiple confidence estimates as indices of eyewitness memory. *Journal of Experimental Psychology: General*, *137*, 528-547.

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 25*, 459-473.

Steblay, N. K., Dysart, J. E., & Wells, G. L. (2011). Seventy-two tests of the sequential lineup superiority effect: A meta-analysis and policy discussion. *Psychology, Public Policy, and Law, 17*, 99-139.

Weber, N., Brewer, N., Wells, G. L., Semmler, C., & Keast, A. (2004). Eyewitness identification accuracy and response latency: The unruly 10-12 second rule. *Journal of Experimental Psychology: Applied, 10*, 139-147.

Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist*, *48*, 553-571.

Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness identification: Improving its probative value. *Psychological Science in the Public Interest*, *7*, 45-75.

Wells, G. L., & Olson, E. A. (2002). Eyewitness identification: Information gain from incriminating and exonerating behaviors. *Journal of Experimental Psychology: Applied, 8*, 155-167.

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 photo spreads. *Law and Human Behavior, 22,* 603-647.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41*, 67-85.

Wickelgren, W. A., & Norman, D. A. (1966). Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology, 3,* 316-347.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114,* 152-176.

Yonelinas A. P., & Jacoby L. L. (1994). Dissociations of processes in recognition memory: Effects of interference and of response speed. *Canadian Journal of Experimental Psychology, 48,* 516-535.

Table 1

Proportion correct decisions in Experiments 1, 2 and 3 for control and deadline confidence conditionswith inferential comparisons for each stimulus event and overall.

	Lineup Condition				
	Control	Deadline Confidence	$\chi^2(1)$	*p*	*w*
Experiment 1					
Stimulus 1	.44	.63	15.17	< .01	0.19
Stimulus 2	.57	.65	2.87	.09	0.08
Overall	.51	.65	17.79	<.01	0.14
Experiment 2					
Stimulus 1	.58	.82	19.68	< .01	0.26
Stimulus 2	.55	.62	1.62	.20	0.07
Stimulus 3	.74	.86	6.66	.01	0.15
Stimulus 4	.50	.64	5.62	.02	0.14
Overall	.59	.73	26.57	< .01	0.15
Experiment 3					
Stimulus 1	.26	.64	17.54	< .001	0.39
Stimulus 2	.36	.51	2.55	.11	0.15
Stimulus 3	.59	.75	3.35	.07	0.17
Stimulus 4	.29	.34	0.28	.59	0.05
Overall	.38	.63	30.76	< .001	0.26

Table 2

Number of decisions, and proportion correct, in Experiments 1, 2 and 3 for varying discrepancies between the maximum (Max) and next highest (next) confidence value.

	Experiment 1		Experiment 2		Experiment 3	
Discrepancy	Proportion Correct	N	Proportion Correct	N	Proportion Correct	N
100	1.00	8	.97	33	1.00	3
≥ 90	1.00	11	.89	46	1.00	5
≥ 80	.94	16	.81	64	.83	6
≥ 70	.90	20	.80	83	.70	10
≥ 60	.70	30	.80	96	.64	14
≥ 50	.64	42	.69	128	.68	19
≥ 40	.55	69	.66	147	.56	25
≥ 30	.51	106	.56	202	.47	38
≥ 20	.44	164	.51	270	.42	57
≥ 10	.37	258	.40	388	.34	104

Supplementary Table

Frequency (and proportion) of lineup classifications in Experiments 1, 2 and 3 for control and deadline confidence conditions by target presence and for each stimulus event and overall.

	Deadline Confidence			Control		
	Suspect ID	Rejection	Inconclusive	Suspect ID	Rejection	Foil ID
			Target Present			
Experiment 1						
Stimulus 1	40	58	2	33	31	37
	(.40)	(.58)	(.02)	(.33)	(.31)	(.37)
Stimulus 2	41	54	5	31	59	14
	(.41)	(.54)	(.05)	(.30)	(.57)	(.14)
Overall[a]	83	112	5	64	90	51
	(.42)	(.56)	(.03)	(.31)	(.44)	(.25)
Experiment 2						
Stimulus 1	52	21	0	44	12	17
	(.71)	(.29)	(.00)	(.60)	(.16)	(.23)
Stimulus 2	24	51	0	21	36	16
	(.32)	(.68)	(.00)	(.29)	(.49)	(.22)
Stimulus 3	56	17	0	53	12	12
	(.77)	(.23)	(.00)	(.69)	(.16)	(.16)
Stimulus 4	22	49	0	20	36	18
	(.31)	(.69)	(.00)	(.27)	(.49)	(.24)
Overall[a]	154	138	0	138	96	63

	(.53)	(.47)	(.00)	(.46)	(.32)	(.21)
Experiment 3						
Stimulus 1	13	14	3	6	6	17
	(.43)	(.47)	(.10)	(.21)	(.21)	(.59)
Stimulus 2	1	29	0	2	17	10
	(.03)	(.97)	(.00)	(.07)	(.59)	(.34)
Stimulus 3	14	13	2	14	5	9
	(.48)	(.45)	(.07)	(.50)	(.18)	(.32)
Stimulus 4	7	5	17	0	13	17
	(.24)	(.17)	(.59)	(.00)	(.43)	(.57)
Overall[a]	35	83	0	22	41	53
	(.30)	(.70)	(.00)	(.19)	(.35)	(.46)
			Target Absent			
Experiment 1						
Stimulus 1	4	94	2	1	54	42
	(.04)	(.94)	(.02)	(.01)	(.56)	(.43)
Stimulus 2	8	84	8	6	92	14
	(.08)	(.84)	(.08)	(.05)	(.82)	(.13)
Overall[a]	16	178	6	7	146	56
	(.08)	(.89)	(.03)	(.03)	(.70)	(.27)
Experiment 2						
Stimulus 1	6	67	0	4	42	30
	(.08)	(.92)	(.00)	(.05)	(.55)	(.39)

Stimulus 2	4	67	0	2	61	13
	(.06)	(.94)	(.00)	(.03)	(.80)	(.17)
Stimulus 3	3	70	0	1	57	13
	(.04)	(.96)	(.00)	(.01)	(.80)	(.18)
Stimulus 4	4	71	0	2	54	18
	(.05)	(.95)	(.00)	(.03)	(.73)	(.24)
Overall[a]	17	275	0	9	214	74
	(.06)	(.94)	(.00)	(.03)	(.72)	(.25)
Experiment 3						
Stimulus 1	3	25	1	0	9	20
	(.10)	(.86)	(.03)	(.00)	(.31)	(.69)
Stimulus 2	0	29	0	0	19	10
	(.00)	(1.00)	(.00)	(.00)	(.66)	(.34)
Stimulus 3	0	30	0	2	20	8
	(.00)	(1.00)	(.00)	(.07)	(.67)	(.27)
Stimulus 4	1	13	16	0	17	11
	(.03)	(.43)	(.53)	(.00)	(.61)	(.39)
Overall[a]	4	114	0	2	65	49
	(.03)	(.97)	(.00)	(.02)	(.56)	(.42)

Note. [a] Overall performance for the deadline confidence condition represents an application of the classification algorithm to the entire data set collapsed across stimuli, not to the sum of the results from each of the separate stimulus sets.