Law and Human Behavior, Vol. 30, No. 1, February 2006 (© 2006)
DOI: 10.1007/s10979-006-9002-7

Eyewitness Identification Accuracy and Response Latency

Neil Brewer,[1,2] Alita Caon,[1] Chelsea Todd,[1] and Nathan Weber[1]

Published online: 28 April 2006

Eyewitness identification research has reliably shown that accurate identifications are faster than inaccurate identifications. Recently, D. Dunning and S. Perretta (2002) claimed that an identification latency of 10–12 s not only best discriminates between accurate and inaccurate identifications but also produces extremely high accuracy rates, approaching 90%. Consistent with predictions from recognition memory theory, however, we show that the optimum time boundary varies with overall response latency under manipulations of retention interval and nominal lineup size, and that the accuracy rate inside the optimum time boundary is much less impressive than previously reported. We outline directions for clarifying the accuracy and latency relationship to assist the reliable diagnosis of identification accuracy.

KEY WORDS: eyewitness identification; identification latency; identification accuracy-latency relationship.

Eyewitness identifications constitute important evidence in many criminal investigations and trials, but are also known to be fallible. It is not surprising, therefore, that a large amount of research attention has been directed at improving the accuracy of eyewitness identifications. This research can be broadly divided into two categories. The first category includes research that aims to improve identification accuracy by improving the lineup procedure itself. Examples of this work (for review, see Brewer, Weber, & Semmler, 2005) include investigations of system variables such as lineup presentation mode (Lindsay & Wells, 1985), lineup fairness (Lindsay & Wells, 1980; Wells, Leippe, & Ostrom, 1979), and the instructions given to a witness before viewing the lineup (Malpass & Devine, 1981). The second category includes research that has focused on the investigation of variables that may allow the identification of either correct or incorrect decisions rather than on improving the accuracy of the identification itself. In other words, researchers have searched for markers of the accuracy of identification decisions (i.e., assessment

[1] Flinders University, Adelaide, South Australia.
[2] To whom correspondence should be addressed at School of Psychology, Flinders University, GPO Box 2100, Adelaide, South Australia 5001, Australia; e-mail: neil.brewer@flinders.edu.au.

variables, Sporer, 1993). One potential marker of identification accuracy that has received considerable research attention is response latency. The present research follows the second of these approaches and investigates the relationship between response latency and accuracy, with a particular focus on the time boundary that best discriminates correct from incorrect decisions under two forensically relevant manipulations, one of retention interval (an estimator variable) and the other of nominal lineup size (a system variable).

A consistent finding in the eyewitness identification literature is a negative identification response latency–accuracy relationship (Brewer, Gordon, & Bond, 2000; Dunning & Perretta, 2002; Dunning & Stern, 1994; Smith, Lindsay, & Pryke, 2000; Smith, Lindsay, Pryke, & Dysart, 2001; Sporer, 1992, 1993, 1994; Weber, Brewer, Wells, Semmler, & Keast, 2004). Importantly, this relationship appears to hold only for witnesses who make a positive identification from the lineup (i.e., choosers) but not for those who reject the lineup (i.e., nonchoosers). Despite the consistency of this relationship, response latency is not by itself a useful marker of identification accuracy in the applied context. The problem arises from the inability of lineup administrators to classify any single identification decision as fast (and, therefore, likely to be accurate) or slow (and, therefore, likely to be inaccurate). The use of discriminant function analysis (Smith et al., 2000; Smith et al., 2001; Sporer, 1994) is one potential solution to this problem. However, the need to calculate a different discriminant function for each set of stimuli and viewing conditions renders this approach unsatisfactory as an applied technique for predicting the accuracy of any single identification decision. An alternative solution is to use time boundary analysis to identify the time boundary that best discriminates correct from incorrect decisions (Dunning & Perretta, 2002). A time boundary analysis is conducted by computing a series of chi-square statistics. Specifically, for each time boundary to be considered, the chi-square statistic for the 2 (time boundary: identification faster than time boundary vs. identification slower than time boundary) $\times$ 2 (accuracy: correct vs. incorrect) contingency table is computed. The time boundary that optimally discriminates correct from incorrect identifications is defined as the time boundary at which the greatest chi-square value is observed.

Dunning and Perretta (2002) used time boundary analyses to identify the optimum time boundary for data from four different experiments. These analyses produced two key findings. First, for all four samples, the optimum time boundary fell within a 10–12-s window. Second, the proportion of correct decisions made before the optimum time boundary was an impressively high 87.1%. On the basis of these results Dunning and Perretta concluded that the optimum time boundary was likely to be constant across stimuli and viewing conditions, and would consistently identify lineup decisions with a high probability of being accurate. Consequently, they suggested the 10–12 s rule for identifying decisions that are likely to be correct. Specifically, they argued that decisions made in 10–12 s or less can be considered likely to be accurate. If such a rule generalized across stimuli and viewing conditions, there would be obvious practical implications for the diagnosis of accurate identification decisions. To investigate the consistency of this result Weber et al. (2004) recently conducted time boundary analyses on data from four experiments, all of which used the same stimulus video. These analyses clearly demonstrated that the time boundary was not consistent across experiments or across the two targets featured in the video. Furthermore, the accuracy rates of decisions made before the optimum time boundary (and also before Dunning and Perretta's suggested 10-s time boundary) were substantially lower than those observed by Dunning and Perretta. Thus, the two key findings from Dunning and Perretta's time boundary analyses, upon which the 10–12-s rule was based, did not generalize to the stimuli and viewing conditions examined by Weber et al. From an applied perspective, Weber et al.'s findings have obvious implications for the utility, or lack thereof, of the 10–12 s rule in the real world. However, their results also raise a number of important theoretical issues regarding the nature of the identification decision process.

Dunning and Perretta (2002) argued that a consistent optimum time boundary accompanied by highly accurate pre-time boundary decisions are predicted by Dunning and Stern's (1994) account of the eyewitness identification decision processes. Specifically, Dunning and Stern described two different decision processes that may be used by witnesses. Automatic decisions are characterized as fast and unconscious and tend to be reported by witnesses who made an accurate identification. In contrast, deliberative decisions, which are more likely to be used by witnesses who make inaccurate identifications, are characterized as slow and thoughtful. Importantly for the time boundary, Dunning and Stern argued that the latency of automatic decisions should be insensitive to changes in the decision context. In other words, changes in stimuli or viewing conditions that influence response latency should affect the latency of deliberative, but not automatic, decisions. Consequently, Dunning and Perretta argued that, regardless of the lineup and viewing conditions, a concentration of fast and accurate decisions should always be present. Therefore, the optimum time boundary should (a) be consistent across viewing conditions and stimuli, and (b) consistently identify decisions with a high probability of being correct. Obviously, Weber et al.'s (2004) demonstration of both the variable time boundary and less than impressive accuracy of pre-time boundary decisions, directly contradicts these predictions. The important issue for understanding the identification decision process, therefore, appears to be why Dunning and Stern's predictions were not supported.

Resolution of this issue is not just of theoretical interest. An understanding of the identification decision process could provide useful practical insights into the ideal process for conducting a lineup (for a broader discussion of this issue, see Brewer, Weber, & Semmler, in press). If Dunning and Stern's (1994) account is correct, then an understanding of the factors that determine the decision process used by witnesses would suggest two potential methods for the improvement of overall identification accuracy. First, the ability to identify decisions made using an automatic decision process would allow lineup administrators to diagnose decisions with a high probability of being correct and, therefore, weight the identification evidence more strongly than a nonautomatic decision. Second, an understanding of the determinants of the decision process used by witnesses may allow the development of a lineup administration or construction method that promotes automatic and, therefore, highly accurate identification decisions. Thus, investigation of the underlying cause of the discrepancy between Dunning and colleagues' results (Dunning & Perretta, 2002) and predictions (Dunning & Stern, 1994) and the results reported by Weber et al. (2004) is an important issue.

One potential account of this difference is that one of the assumptions underlying the automatic–deliberative processing account of eyewitness identification decision-making is incorrect. Specifically, consideration of the broader cognitive psychology literature suggests that the invariance of the response latency of automatic decisions posited by Dunning and colleagues (Dunning & Perretta, 2002; Dunning & Stern, 1994) is questionable. For example, theories of recognition memory based on information accumulation (Van Zandt, 2000) and the diffusion process (Ratcliff, 1978) predict the difference in correct and incorrect decision latencies observed in the eyewitness identification domain. These theories also predict that response latency will be affected by manipulations that influence the discriminability of old from new stimuli. Both theories posit that a decision is made when the evidence accumulated by an individual reaches a criterion value. Importantly, as discriminability decreases the rate of accumulation of clear-cut evidence slows and decisions take, on average, longer. Importantly, this change in response latency is predicted to occur for both correct and incorrect decisions. Therefore, theories of recognition memory that include predictions relevant to response latency suggest that the optimum time boundary is likely to be variable. Specifically, they predict that the optimum time boundary should be affected by manipulations of discriminability that affect overall response latency. This prediction is consistent with the results of Weber et al. (2004) who observed that, in the data sets they analyzed, the optimum time boundary tended to vary with overall response latency. Therefore, a primary aim of the experiments reported here was to test the competing predictions of the automatic–deliberative account of eyewitness identification decision processes (Dunning & Perretta, 2002; Dunning & Stern, 1994) and recognition memory theories (Ratcliff, 1978; Van Zandt, 2000). Specifically, we investigated the impact of two forensically relevant manipulations (i.e., retention interval in Experiment 1 and nominal lineup size in Experiment 2) on overall response latency and on the optimum time boundary. We did so using encoding and test stimuli not previously used by Dunning and colleagues or Weber et al. (2004).

EXPERIMENT 1

In Experiment 1 we examined how the optimum time boundary varied with retention interval, that is, the interval between the viewing of the target stimulus and the identification test. Recognition memory theories have long considered response latency to be a key indicator of underlying memory strength (Kahana & Loftus, 1999; Murdock & Dufty, 1972) and, hence, of the likelihood that a previously studied stimulus is recognized as having been seen before (Atkinson & Juola, 1974; Gillund & Shiffrin, 1984). Memory strength is, in turn, considered to be affected by variables such as retention interval (Metcalfe, 1996; Murdock, 1985). To the extent that increasing the retention interval undermines memory quality for the target stimulus, correct and incorrect identification latency should be increased, as should the optimum time boundary. Three retention intervals—0, 15, and 30 min—were examined.

Methods

Participants

Two hundred and thirty-eight individuals (93 male, 145 female), completed this experiment. Participants were either first-year psychology students, who participated as part of a research participation exercise, paid volunteers recruited on campus, or unpaid volunteers recruited from the community.

Materials

A single video of a nonviolent, staged crime was presented to all participants. It featured a white female shoplifting from a supermarket. The shoplifter walked along a supermarket aisle, then removed an item from the shelves and examined it before slipping it into her handbag and walking away from the camera. The film lasted for 42 s, with the shoplifter's face in view for approximately 6 s. A photograph of the offender for use in the target-present lineups, was taken after filming the video. The offender wore different clothes in the video and lineup photograph. Further, the photograph of the shoplifter depicted her with hair down, rather than tied back as in the video. Eight foils were selected from a pool of photographs on the basis of their match to the description of the offender. One of these eight foils was randomly chosen as the target's replacement for the target-absent lineup. All stimuli were presented on a 17″ PC monitor with resolution set at $1{,}024 \times 768$ pixels. The video was displayed at a size of 670×500 pixels. The lineups were presented as two rows of four photographs, each presented at a size of 200×200 pixels (4.5-cm square). A button labeled *not present* was presented below the photo array. The arrangement of photographs in the array was randomly determined for each participant.

Design and Procedure

To examine the impact of retention interval on predictors of identification accuracy, a 3 (retention interval: 0, 15, and 30 min) $\times$ 2 (target presence: target-present and target-absent) between-participants design was used. With the exception of the distractor task, the experiment was completed on computer with each participant in an individual cubicle. Before watching the stimulus video all participants were informed that they would be asked to watch a video of a staged crime and make an identification decision about the person or people in it. After viewing the video, participants in the 15 or 30 min retention interval completed the distractor task (picture puzzles) for the allotted time. Participants in the 0 min condition proceeded immediately to the identification test.

Before being presented with the lineup, participants were explicitly informed that (a) the offender may or may not be present in the photo array and (b) if they thought the offender was present, they should click on that offender's photograph, or if they thought the offender was not present, they should click on the *not present* button. The lineup was then presented on the screen and remained in view until the participant had indicated their decision by clicking on a photograph or the *not*

Table 1. Frequency and Percentage of Identification Responses in Each Condition for Target-Present and Target-Absent Lineups

	Correct identification		Incorrect identification		Rejection		
Condition	No	%	No	%	No	%	Total No
Target-present							
0 min	12	30.0	7	17.5	21	52.5	40
15 min	13	33.3	7	17.9	19	48.7	39
30 min	12	30.0	10	25.0	18	45.0	40
Overall	37	31.1	24	20.2	58	48.7	119
Target-absent							
0 min			18	46.2	21	53.8	39
15 min			16	40.0	24	60.0	40
30 min			19	47.5	21	52.5	40
Overall			53	44.5	66	55.5	119

present button. Response latency was recorded by the computer as the time from the onset of the lineup display to the participant's response.

Results

An alpha level of $\alpha = .05$ was used for all inferential analyses. Cohen's[3] f is reported as the measure of effect size for ANOVAs and Cohen's w (which is equivalent to the phi coefficient for 2×2 contingency tables) is reported for chi-square analyses. The frequencies of the different categories of identification response, for both target-present and -absent lineups, are displayed in Table 1. Chi-square analyses were used to examine the effect of the manipulation on identification responses. For target-present lineups, a 3 (identification response) $\times$ 3 (retention interval) chi-square analysis revealed no significant impact of retention interval on the frequency of the different identification responses, $\chi^2(4, n = 119) = 1.02$, $w = 0.09$. Similarly, for target-absent lineups a 2 (identification response) $\times$ 3 (retention interval) chi-square analysis revealed no significant impact of retention interval on identification responses, $\chi^2(2, n = 119) = 0.52$, $w = 0.07$.

Following previous investigation of response latency in eyewitness identification (e.g., Sporer, 1992), separate analyses were conducted for choosers and nonchoosers. Descriptive statistics for response latency of choosers and nonchoosers by accuracy and retention interval are presented in Table 2. Because of significantly positively skewed response latency distributions in all conditions, inferential analyses were conducted on transformed data (i.e., log base 10). As the transformed and nontransformed data displayed the same pattern of results, descriptive statistics are based on the nontransformed data to assist interpretation. For both choosers and nonchoosers a 3 (retention interval) $\times$ 2 (accuracy) ANOVA was conducted on transformed response latency. Consistent with previous findings, correct identifications were made, on average, faster than incorrect identifications,

[3]Values of Cohen's f greater than .4 are considered large effects, while the cut-off values for small and medium effects are .1 and .25, respectively.

Table 2. Descriptive Statistics for Response Latency of Choosers and Nonchoosers by Accuracy and Retention Interval

	Accurate			Inaccurate			Overall		
Choice	*M*	*SD*	95% CI	*M*	*SD*	95% CI	*M*	*SD*	95% CI
Choosers									
0 min	16.98	10.16	10.52–23.43	22.13	11.42	17.42–26.83	20.46	11.16	16.74–24.18
15 min	25.96	14.06	17.46–34.46	34.66	18.62	26.61–42.71	31.52	17.43	25.62–37.41
30 min	22.51	16.45	12.06–32.96	36.42	45.61	19.07–53.76	32.35	39.64	19.83–44.86
Overall	21.93	13.94	17.28–26.58	31.25	30.80	24.26–38.24	28.23	26.82	23.25–33.20
Nonchoosers									
0 min	25.53	15.10	18.66–32.41	18.54	6.50	15.59–21.50	22.04	12.01	18.29–25.78
15 min	37.86	31.11	24.72–50.99	26.23	15.07	18.97–33.50	32.72	25.72	24.81–40.64
30 min	23.80	18.09	15.56–32.03	23.69	13.40	17.03–30.35	23.75	15.89	18.59–28.90
Overall	29.46	23.55	23.67–35.25	22.66	12.28	19.43–25.89	26.28	19.36	22.84–29.72

$F(1, 108) = 4.97$, $f = 0.21$, but no main effect of accuracy was observed for nonchoosers, $F(1, 118) = 0.00$, $f = 0.00$. Similarly, a significant main effect of retention interval was identified for choosers, $F(2, 108) = 4.07$, $f = 0.27$, but not nonchoosers, $F(2, 118) = 1.91$, $f = 0.18$. Examination of the descriptive statistics suggests that (a) correct identifications were made, on average, faster than incorrect identifications, and (b) identifications in the 0 min condition were made faster than those in the 15 and 30 min conditions. No significant interaction effect was found for choosers, $F(2, 108) = 0.02$, $f = 0.02$, or nonchoosers, $F(2, 118) = 0.25$, $f = 0.07$. In other words, there is no evidence that the retention interval manipulation affected the difference in response latency between correct and incorrect identifications.

As no evidence of a significant response latency–accuracy relationship was present for lineup rejections, time boundary analyses were performed only on the choosers' data. Time boundary analyses were conducted separately for each retention interval. Consistent with the technique used by Dunning and Perretta (2002) and Weber et al. (2004), a chi-square statistic[4] based on the 2 (time boundary: faster or equal vs. slower) $\times$ 2 (accuracy: correct vs. incorrect) contingency tables was computed with the time boundary set at each integer value from 1 s to 50 s (i.e., 1 s, 2 s, 3 s, and so on until 50 s). The latency that produced the greatest chi-square value (i.e., the highest peak in the time boundary curve) was identified as the time boundary that optimally discriminated correct from incorrect decisions. Following Weber et al., we also report the confidence range for each peak in the time boundary curve. The confidence range includes all consecutive time boundaries that produce a chi-square value within ± 1 standard error of the peak value. The confidence range is indicative of the degree to which the peak can be confidently located at a specific time boundary.

Plots of chi-square value (and standard error) by time boundary for each retention interval are presented in Fig. 1. Consistent with the main effect of retention interval on mean response latency for choosers, the optimal time boundary also appears to differ between retention interval conditions. Specifically, the optimal time boundary for the 0 min condition (13 s, $w = 0.31$, confidence range = 13–14 s) is markedly earlier than those for the 15 min (36 s, $w = 0.34$, confidence range = 36–42 s) and 30 min (35 s, $w = 0.27$, confidence range = 35–39 s) conditions. Thus, the time boundary appears to be later in the conditions that produced, on average, slower identification decisions. This observation parallels the findings of Weber et al. who demonstrated that, in the data sets they analyzed, the optimal time boundary generally varied with mean response latency.

In addition to the variability of the optimal time boundary itself, the other striking features of these data are the proportions of correct identifications made before the optimum time boundaries (0 min.: 54.5%; 15 min.: 46.2%; 30 min: 36.7). In stark contrast to the observation by Dunning and Perretta (2002) of almost 90% correct identifications made before the optimum time boundary, none of these conditions produced pretime boundary accuracy rates in excess of 60%.

[4]Time boundary analyses can also be conducted by computing the log odds ratio at each time boundary. As the log odds ratio analyses produce the same pattern of results as the chi square analyses, we report only the chi square analyses.

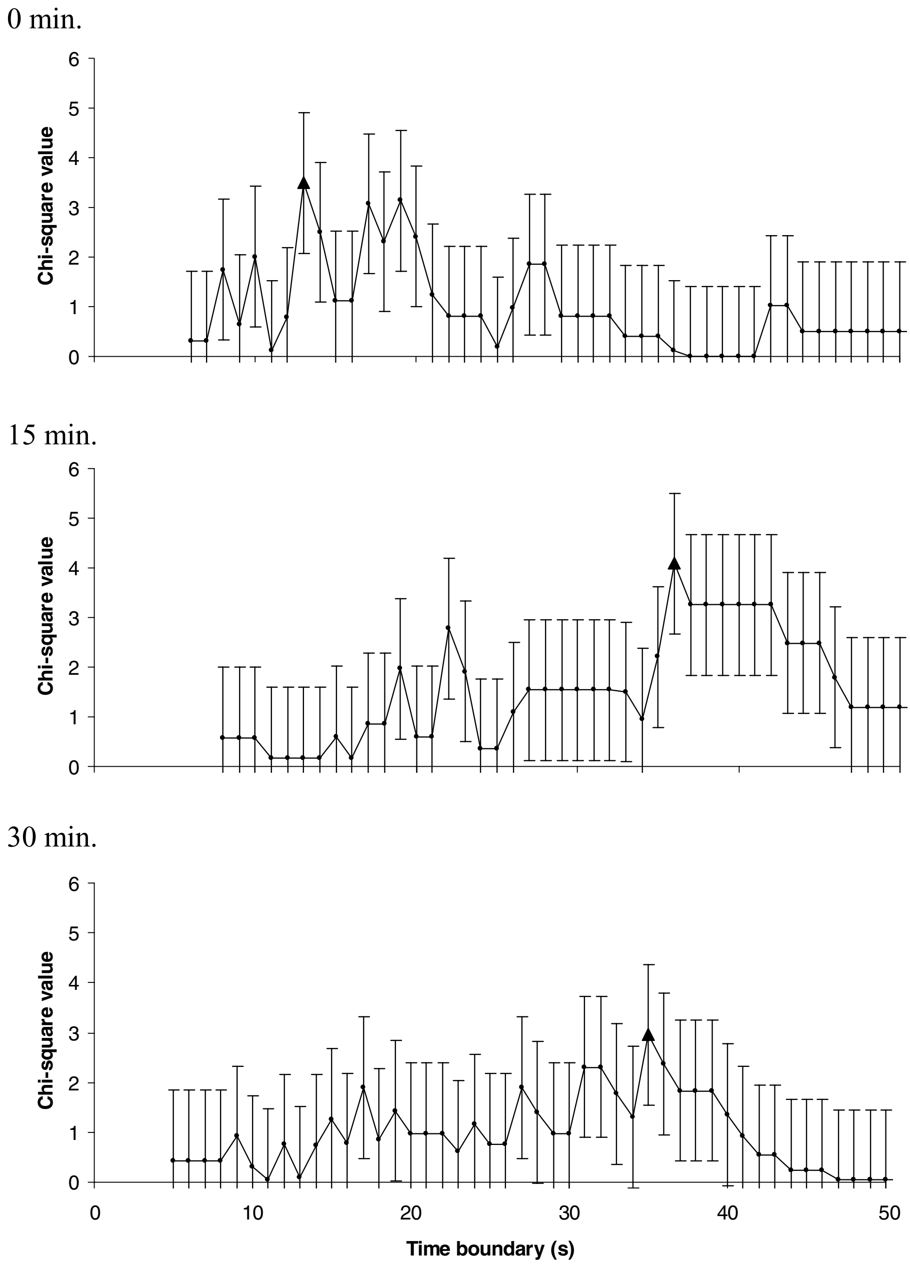

Fig. 1. Plots of chi-square (and standard error) by time-boundary for each retention interval condition.

Discussion

In sum, these data replicated the well established negative response latency–accuracy relationship for choosers. Furthermore, for choosers, the manipulation of retention interval was demonstrated to influence both mean response latency and the latency that optimally discriminated correct from incorrect identifications. Of course, the retention intervals used here (i.e., 0, 15, and 30 min) are negligible in comparison with the days, weeks, or months that can pass between the crime and a witness viewing a lineup in the real world. However, the observed impact of such a seemingly trivial manipulation provides a particularly strong demonstration of the instability of the optimum time boundary. Notably, in all conditions, the optimum time boundary was outside Dunning and Perretta's (2002) 10–12-s window. Despite the variability in the optimum time boundary, the proportion of correct decisions made before the boundary was consistently poor. Finally, the impact of the retention interval manipulation on the optimum time boundary has important implications for our understanding of the identification decision process.

One potential explanation of the observed variability in the optimum time boundary between conditions is the relatively small number of choosers in each condition (i.e., 37, 36, and 41 for the 0-, 15-, and 30-min retention intervals, respectively). Although this may account for our failure to replicate the 10–12 s window, the small number of choosers does not explain the observation of markedly later optimal time boundaries in conditions with longer average response latencies. Two theoretical explanations of the impact of the manipulation on the optimum time boundary appear plausible. First, as predicted by the recognition memory theories discussed in detail earlier (Ratcliff, 1978; Van Zandt, 2000), the latency of all decisions, including those based on an automatic decision process, is susceptible to manipulations of discriminability. An alternative explanation, however, is that these stimuli did not elicit the "significant plurality" (Dunning & Perretta, 2002, p. 960) of participants relying on an automatic decision process to produce a stable time boundary, a possibility that is consistent with the low accuracy rates observed for target-present lineups. Consequently, the time boundaries were determined largely by the unstable response latency of deliberative decisions. The implications of these accounts are addressed in detail in the General Discussion section.

EXPERIMENT 2

The results of Experiment 1 provided a clear demonstration of the impact of changes in the retention interval on the optimum time boundary. In Experiment 2 we aimed to provide converging evidence for the instability of the optimum time boundary by manipulating another forensically relevant factor likely to affect discriminability. Rather than manipulate another factor likely to influence the quality of the memory of the offender held by the participants, we manipulated the conditions under which participants were asked to identify the offender from a lineup, specifically, varying the nominal size of the lineup. Such a manipulation provides a particularly strong test of the stability of the optimum time boundary as Dunning and Perretta (2002) argued that such a manipulation should not affect the latency of automatic decisions or the optimum time boundary.

Methods

Participants

One hundred and ninety-six individuals (33 male, 163 female), completed this experiment. Participants were paid volunteers recruited from the community (i.e., staff, parents, and friends) of a local school.

Materials

In addition to the video used in Experiment 1, a second video of a nonviolent staged crime was also used in this experiment. The second video displayed a White male attempting to break into a car. Before successfully opening the car door he was startled by a shout and ran from the scene. The video lasted for 16 s and the car thief's face was in view for approximately 3 s. The lineup photograph of the car thief was taken after filming the video. The car thief was wearing different clothes, but his appearance was otherwise unaltered. Both videos were presented, as in Experiment 1, at a size of 670 × 500 pixels, on a monitor with resolution set at 1,024 × 728 pixels.

For the car thief lineup, 12 foils were chosen from a pool of photographs based on their match to the description of the offender and one was chosen as the target's replacement for the target-absent lineups. For the shoplifter lineup, in addition to those used in Experiment 1, four foils were chosen based on their match to the description of the offender. The same photograph was used as the target's replacement in both experiments. Depending on the lineup size condition, the lineups were presented as one, two, or three rows of four photographs, each presented at a size of 200 × 200 pixels. As in Experiment 1, a button labeled *not present* was presented below the photoarray and the arrangement of photographs in the array was randomly assigned for each participant. Further, given this random positioning, the onscreen positions of the rows of photographs, combined with the position of the cursor at lineup onset (a common point at the bottom of the screen), ensured that the average mouse travel distance (across participants) was equivalent across lineup size conditions. Thus, the manipulation of nominal lineup size was not confounded with the average distance mouse movement required to indicate a decision.

Design and Procedure

To examine the impact of lineup size on the predictors of identification accuracy, a 3 (lineup size: four, eight, and 12) × 2 (target presence: target-present and target-absent) between-subjects design was used. For all participants the shoplifter video was viewed and a decision made about the shoplifter lineup before presentation of the car thief video and lineup. For both stimuli, after viewing the video participants were immediately shown the lineup instructions and asked to make a

Table 3. Frequency and Percentage of Identification Responses in Each Condition for Target-Present and Target-Absent Lineups for Both Videos

Conditions	Correct identification No	Correct identification %	Incorrect identification No	Incorrect identification %	Rejection No	Rejection %	Total No
Shoplifter Video							
Target-present							
4	13	40.6	2	6.3	17	53.1	32
8	7	21.9	7	21.9	18	56.3	32
12	7	20.6	8	23.5	19	55.9	34
4							
Overall	27	27.6	17	17.3	54	55.1	98
Target-absent							
4			9	28.1	23	71.9	32
8			12	35.3	22	64.7	34
12			12	37.5	20	62.5	32
Overall			33	33.7	65	66.3	98
Car thief video							
Target-present							
4	26	81.3	4	12.5	2	6.3	32
8	22	64.7	4	11.8	8	23.5	34
12	20	62.5	5	15.6	7	21.9	32
Overall	68	69.4	13	13.3	17	17.3	98
Target-absent							
4			16	50.0	16	50.0	32
8			22	68.8	10	31.3	32
12			25	73.5	9	26.5	34
Overall			63	64.3	35	35.7	98

decision from the lineup. The lineup instructions were identical to those used in Experiment 1. Participants viewed one of three different sized photoarrays: 4-, 8-, or 12-person lineups. In the four- and eight-person lineup conditions, for each participant the foils presented with the target or target's replacement were randomly selected (by the identification test software) from the pool of 11 foils. For the 12-person lineup condition all 11 foils were presented.

Results

Table 3 displays the frequencies of the different categories of identification response, for both target-present and -absent lineups. Chi-square analyses were used to examine the effect of the manipulation on identification responses. For target-present lineups, 3 (identification response) $\times$ 3 (lineup size) chi-square analyses revealed no significant impact of lineup size on the frequency of identification responses for either the shoplifter video, $\chi^2(4, n=98)=6.39$, $w=0.25$, or the car thief video, $\chi^2(4, n=98)=4.58$, $w=0.22$. Similarly, for target-absent lineups 2 (identification response) $\times$ 3 (lineup size) chi-square analyses revealed no significant impact of lineup size on identification responses for either the shoplifter, $\chi^2(2, n=98)=0.69$, $w=0.08$, or car thief, χ^2 $(2, n=98)=4.39$, $w=0.21$, videos. Examination of the effect size measures, however, suggests small effects were present for all but the target-absent lineups for the shoplifter video.

Table 4. Descriptive Statistics for Response Latency of Choosers and Nonchoosers by Accuracy and Lineup Size for Both Videos

	Accurate			Inaccurate			Overall		
Choice	*M*	*SD*	95% CI	*M*	*SD*	95% CI	*M*	*SD*	95% CI
Shoplifter video									
Choosers									
4	11.36	6.99	7.14–15.59	15.89	8.57	10.13–21.65	13.44	7.92	10.09–16.78
8	12.36	4.94	7.79–16.92	25.83	15.56	18.33–33.33	22.20	14.74	16.25–28.16
12	18.08	10.79	8.10–28.06	31.36	19.39	22.28–40.43	27.91	18.35	20.66–35.17
Overall	13.36	7.96	10.21–16.51	25.85	16.86	21.06–30.65	21.47	15.52	17.95–25.00
Nonchoosers									
4	17.72	15.91	10.84–24.59	9.85	5.41	7.07–12.64	14.37	13.05	10.20–18.55
8	19.26	14.82	12.69–25.83	16.87	6.44	13.67–20.07	18.19	11.74	14.43–21.94
12	24.86	16.49	17.15–32.58	21.82	10.70	16.66–26.97	23.38	13.88	18.88–27.88
Overall	20.44	15.78	16.53–24.35	16.40	9.24	13.88–18.92	18.61	13.32	16.19–21.03
Car thief video									
Choosers									
4	8.78	6.10	6.32–11.24	10.35	5.40	7.82–12.88	9.46	5.80	7.74–11.18
8	17.02	11.35	11.99–22.05	19.10	12.26	14.15–24.05	18.15	11.77	14.73–21.57
12	18.49	16.48	10.77–26.20	32.82	21.47	24.80–40.84	27.09	20.70	21.20–32.97
Overall	14.30	12.28	11.33–17.27	22.21	17.93	18.12–26.31	18.48	15.97	15.85–21.11
Nonchoosers									
4	14.07	8.65	9.46–18.68	10.00	3.07	0.00–37.55	13.62	8.26	9.51–17.73
8	21.16	14.89	10.51–31.81	28.97	23.15	9.61–48.32	24.63	18.82	15.27–33.99
12	29.07	20.74	13.13–45.01	26.52	14.04	13.53–39.50	27.96	17.60	18.57–37.34
Overall	19.95	15.22	14.72–25.18	25.73	18.59	16.17–35.28	21.84	16.44	17.26–26.42

Descriptive statistics for the response latency of choosers and nonchoosers by accuracy and lineup size are presented in Table 4. Because of significantly positively skewed response latency distributions in all conditions, inferential analyses were conducted on transformed data (i.e., log base 10). As the nontransformed data displayed the same pattern of results, descriptive statistics are again based on the nontransformed data. For both choosers and nonchoosers 3 (lineup size) $\times$ 2 (accuracy) ANOVAs were conducted, separately for each video, on transformed response latency. For both shoplifter, $F(1, 71) = 12.12$, $f = 0.41$, and car thief videos, $F(1, 138) = 8.46$, $f = 0.25$, the main effect of accuracy was significant, and at least moderately strong, for choosers. However, for nonchoosers the effect was nonsignificant for both videos, shoplifter: $F(1, 113) = 0.62$, $f = 0.02$; car thief: $F(1, 46) = 0.05$, $f = 0.03$. Examination of the descriptive statistics clearly indicates that, for both videos, correct identifications were made faster than incorrect identifications. Similarly, for choosers in both videos a strong and significant main effect of lineup size was observed, shoplifter: $F(2, 71) = 4.86$, $f = 0.37$; car thief: $F(2, 138) = 20.06$, $f = 0.54$. Examination of the descriptive statistics suggests that for the shoplifter video identifications from the four-person lineups were made faster than those from the 8- and 12-person lineups. For the car thief video, identifications from the four-person lineup were made faster than those from the eight-person lineup, which were, in turn, made faster than those from the 12-person lineups. The interaction effect was nonsignificant for choosers from both videos, shoplifter: $F(2, 71) = 0.41$, $f = 0.11$; car thief: $F(2, 138) = 1.62$, $f = 0.15$, although the effect size measures suggest that, with more power, reliable effects may have been observed.

Using the same method as Experiment 1, time boundary curves were plotted for each lineup size for the shoplifter video and car thief videos (Figs. 2 and 3, respectively). Again, because of the nonsignificant main effects of accuracy on response latency for nonchoosers, only choosers' data were used in the time boundary analyses. For the car thief data, a clear impact of lineup size on the optimum time boundary is evident in the time boundary curves. Consistent with the results of Experiment 1, the optimum time boundary appears to vary with mean response latency. Specifically, the four-person lineups, which produced the fastest identification decisions, displayed the smallest optimum time boundary (5 s, $w = 0.36$, confidence range = 5 s). Further, the eight-person lineups produced both intermediate response latency and an intermediate optimum time boundary (8 s, $w = 0.28$, confidence range = 8 s). Finally, the slowest identifications and longest optimum time boundary were produced in the 12-person lineup conditions (23 s, $w = 0.45$, confidence range = 20–28 s). A similar increase is evident for the shoplifter video from the eight-person (10 s, $w = 0.60$, confidence range = 10 s) to the 12-person lineup condition (15 s, $w = 0.48$, confidence range = 15–16 s). However, no clear optimum time boundary is evident at all in the four-person lineup condition for the shoplifter video. The greatest chi-square value was observed at latencies of 16 s through 21 s ($w = 0.35$), but a secondary, and almost equivalent, peak was also observed at 10 s. The lack of a clear time boundary in this condition is emphasized by the confidence range for the greatest chi-square value (confidence

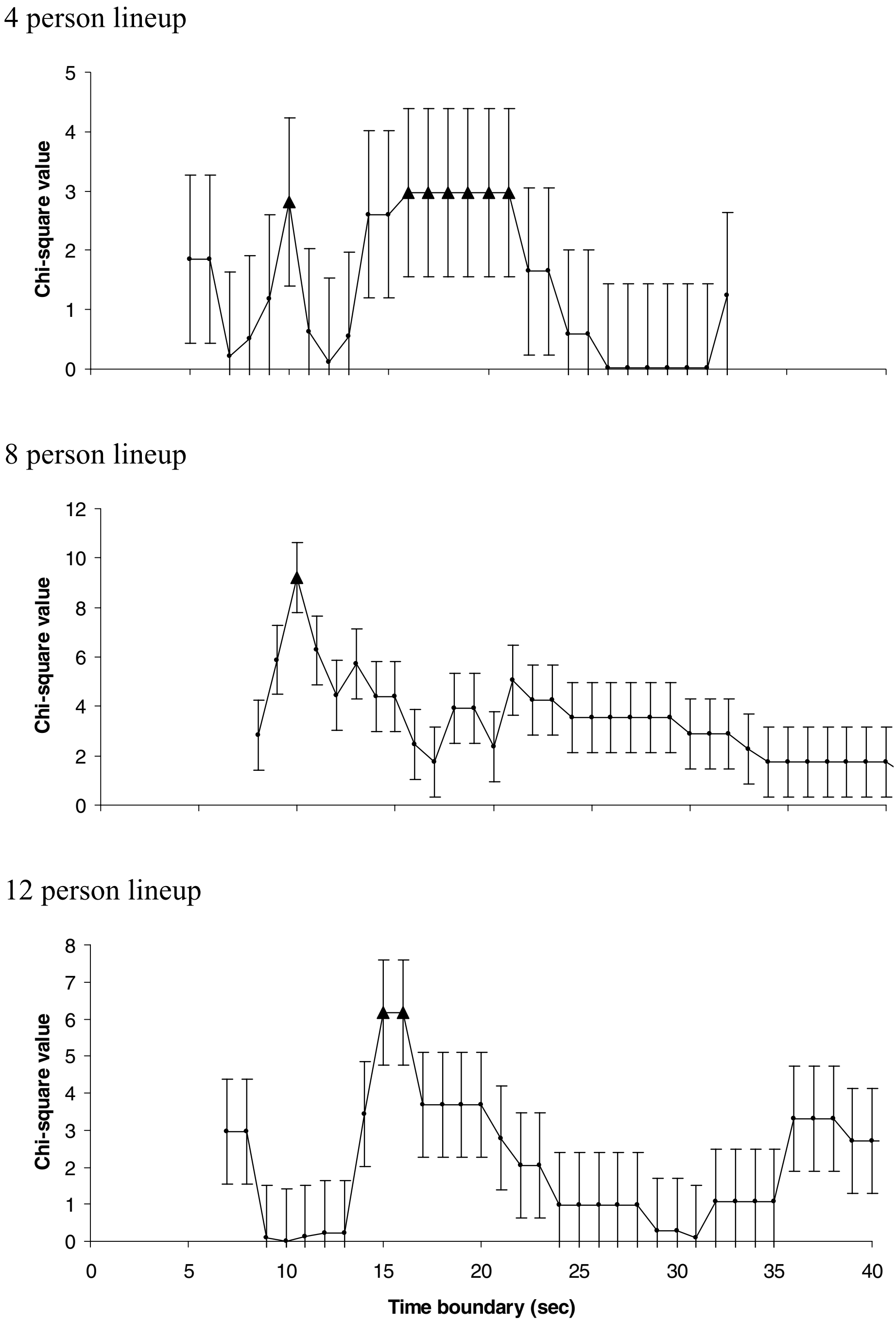

Fig. 2. Plots of chi-square (and standard error) by time-boundary for the shoplifter data in each lineup size condition. Note that to allow easy discrimination of the chi-square peak or peaks for each curve the *y*-axis scales are not consistent.

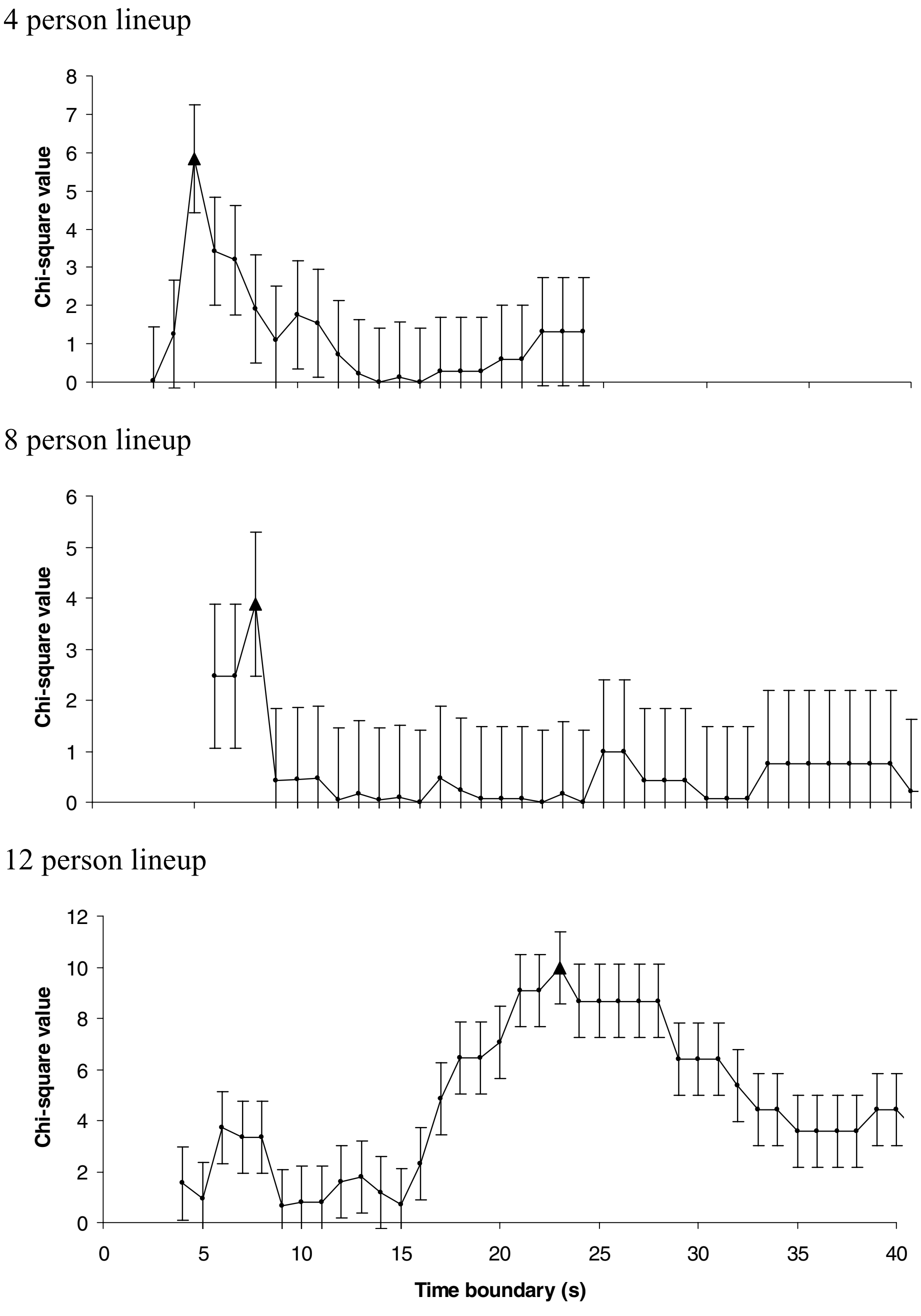

Fig. 3. Plots of chi-square (and standard error) by time-boundary for the car thief data in each lineup size condition. Note that to allow easy discrimination of the chi-square peak or peaks for each curve the y-axis scales are not consistent.

range = 14–23 s). Therefore, with the exception of the four-person shoplifter lineups for which no clear optimum time boundary was observed, these findings parallel those of Experiment 1 and Weber et al. (2004). Specifically, later optimum time boundaries were observed in conditions with slower mean response latency.

As in Experiment 1, the accuracy of decisions made before the optimal time boundary was also considered. In contrast with Experiment 1, high pre-time boundary accuracy rates were evident in a number of conditions. Specifically, accuracy rates above 80% were observed for eight-person lineups from both videos, as well as for the four-person lineups from the car thief video. The remaining conditions evidenced pre-time boundary accuracy rates less than 65%.

Discussion

This experiment provided more evidence of the instability of the optimum time boundary with another new set of stimulus materials. Specifically, the optimum time boundary increased with lineup size and mean response latency. The consistency of this finding (at least in conditions where an optimum time boundary was identifiable) across two sets of stimulus materials, which differed markedly in difficulty, provides a particularly strong demonstration that this variability is not the product of either a particularly easy or difficult lineup. Thus, the conclusion that neither the 10–12 s window or the optimum time boundary are satisfactory markers of identification accuracy in the applied context is further supported.

Again the observation of increased overall latency with later optimum time boundary is consistent with recognition memory theories (Ratcliff, 1978; Van Zandt, 2000). Additionally, as in Experiment 1, the observed variability in the optimum time boundary could be due to a low proportion of automatic decisions. Although this explanation is consistent with the low accuracy observed for shoplifter lineups, given the high accuracy rates for the car thief lineup, this explanation seems unsatisfactory. An alternative explanation is that the participants who reached their decision automatically may not have responded immediately. In other words, after reaching a rapid, automatic decision some participants may have continued to examine the lineup. As the both the average duration of and the likelihood of engaging in this supplementary examination are plausibly associated with the nominal size of the lineup, this delayed responding may account for the observed association between lineup size and the optimum time boundary.

GENERAL DISCUSSION

Two important findings regarding the response latency–accuracy relationship emerged from these experiments. First, using two novel sets of stimulus materials and two forensically relevant manipulations we found clear evidence that the time boundary that best discriminates correct from incorrect identifications is not invariant across lineups or stimulus viewing conditions. Indeed, the fact that the optimum time boundary varied from 5 to 36 s is clear evidence that a simple 10–12 s rule is not sufficient to diagnose correct or incorrect identification decisions. Thus, these results suggest that, although time boundary analysis is a useful research tool, it is not a practically useful marker of identification accuracy. Second, consistent with the findings of Weber et al. (2004) the optimum time boundary was observed to vary with overall response latency. It is significant that the changes in both overall response latency and the optimum time boundary were produced by manipulation of retention interval and lineup size. The automatic–deliberative process account of eyewitness identification decision making (Dunning & Perretta, 2002; Dunning & Stern, 1994) predicts that the latency of deliberative, but not automatic decisions, will be influenced by such manipulations. As this account posits that the optimum time boundary is determined by the latency of automatic decisions, it predicts that these manipulations should influence overall latency, but not the optimum time boundary. Therefore, our findings are consistent with the predictions of recognition memory models (Ratcliff, 1978; Van Zandt, 2000), but apparently inconsistent with the automatic–deliberative account of the eyewitness identification decision process.

One potential account of the results, consistent with Dunning and Stern's (1994) theory, is that the instability of the optimum time boundary is the result of variation in the proportion of automatic decisions. Specifically, Dunning and Perretta (2002) argued that their observation of a stable optimum time boundary was potentially caused by the presence of a significant proportion of fast and accurate automatic decisions. Based on this argument, the stability of the optimum time boundary would be predicted to depend on the proportion of automatic decisions. Thus, the observed variability in optimum time boundary could be the result of a low overall proportion of automatic decisions or a change in the proportion of automatic decisions as a result of the manipulation. An understanding of the factors that influence the relative proportions of automatic and deliberative decisions is, therefore, an important step towards understanding the variables that influence the optimum time boundary. Consideration of the work of Treisman and colleagues (Treisman & Gelade, 1980; Treisman & Gormican, 1988) in the area of visual search suggests a mechanism by which the proportion of automatic and deliberative decisions may be influenced. They argued that when the target in a visual search task can be discriminated from the distractor stimuli by a single distinguishing feature, the target will be identified rapidly, as the result of a pre-attentive, unconscious search of the stimulus array. Therefore, it seems reasonable to suggest that when a witness with a good memory of the offender is presented with a target-present simultaneous lineup, their attention may be drawn unconsciously to the target. If the match between the target and the witness's memory is close enough, as it is likely to be if the witness has a good memory, the witness may then identify the target as the offender without consideration of the remaining lineup members. This mechanism would seem to produce the same kind of rapid, unconscious, and likely to be accurate decisions that Dunning and Stern (1994) characterize as automatic. Importantly, for this mechanism to operate, the target *must* be distinguishable from the foils by some feature or featural configuration (features that are not necessarily describable by the witness). Consequently, in situations where the target does not possess a unique feature or featural configuration or where that feature is not present in the witness's memory of the offender, rapid identification of the target without consideration of the foils is unlikely. Such a mechanism could account for our findings. Specifically, the variability in the time boundary and accuracy rate of choosers for the car thief and overall data from Experiment 2 could be due to a reduction in the proportion of automatic decisions with greater lineup sizes. In contrast, the consistent accuracy and the variable time boundary observed for the shoplifter data could be attributed to a consistently small proportion of automatic decisions across conditions, accompanied by slower deliberative decisions for longer retention intervals and greater lineup sizes. As the car thief stimuli produced higher accuracy rates for choosers than the shoplifter stimuli, this account appears plausible. Importantly, this mechanism also suggests a potential explanation for the discrepancy between the results of Dunning and Perretta (2002) and those found here and by Weber et al. (2004).

Another plausible explanation of these findings is that the automatic–deliberative processes account of the eyewitness identification decision process (Dunning & Perretta, 2002; Dunning & Stern, 1994) is fundamentally flawed. Specifically, the response latency of automatic decisions is not invariant across stimuli and viewing conditions. As already discussed, this idea is consistent with the predictions of basic recognition memory theories (Ratcliff, 1978; Van Zandt, 2000). These theories predict that manipulations of the discriminability of old from new stimuli will affect the response latency of both correct and incorrect decisions. Consistent with the theories, the optimum time boundary was observed to vary with overall latency. As this variation was observed in situations where the accuracy of identification decisions was not affected (i.e., for the shoplifter data in both experiments), this suggests that response latency, and the optimum time boundary, are more sensitive indicators of discriminability than accuracy itself.

These experiments provide further support for Weber et al.'s (2004) conclusion that the optimum time boundary is not invariant and that the 10–12 s rule is not a valid method for the identification of accurate decisions and is, therefore, not suitable for use in an applied setting. Furthermore, these findings highlight two issues that are likely to be important in furthering our understanding of the eyewitness identification process. One is the determinant of witnesses' attention to specific stimuli in the lineup and the impact of this attention on the decision process; the other is the time course of the identification decision process.

In summary, our results provide further evidence for the robust relationship between identification accuracy and response latency. However, they also demonstrate that our understanding of this relationship is as yet insufficient to allow the reliable diagnosis of identification accuracy in the applied context. Perhaps most importantly, however, these findings (a) underscore the need for a comprehensive theoretical approach to the investigation of the eyewitness identification decision process, and (b) demonstrate that a sensible basis for such a theory is provided by work from basic cognition research.

ACKNOWLEDGMENTS

This research was supported by grants A00104516 and DP0556876 from the Australian Research Council.

REFERENCES

Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology: Vol. 1. Learning, memory and thinking* (pp. 243–293). San Francisco, CA: Freeman.

Brewer, N., Gordon, M., & Bond, N. (2000). Effect of photoarray exposure duration on eyewitness identification accuracy and processing strategy. *Psychology, Crime & Law*, *6*, 21–32.

Brewer, N., Weber, N., & Semmler, C. (2005). Eyewitness identification. In N. Brewer, & K. D. Williams (Eds.), *Psychology and law: An empirical perspective* (pp. 177–221). New York: Guilford.

Brewer, N., Weber, N., & Semmler, C. (in press). A role for theory in eyewitness identification research. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. Toglia (Eds.), *Handbook of eyewitness psychology: Volume 2: Memory for people*. Mahwah, NJ: Erlbaum.

Dunning, D., & Perretta, S. (2002). Automaticity and eyewitness accuracy: A 10- to 12-second rule for distinguishing accurate from inaccurate positive identifications. *Journal of Applied Psychology*, *87*, 951–962.

Dunning, D., & Stern, L. B. (1994). Distinguishing accurate from inaccurate eyewitness identifications via inquiries about decision processes. *Journal of Personality and Social Psychology*, *67*, 818–835.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*, 1–67.

Kahana, M., & Loftus, G. (1999). Response time versus accuracy in human memory. In R. J. Sternberg (Ed.), *The nature of cognition* (pp. 323–384). Cambridge, MA: MIT Press.

Lindsay, R. C. L., & Wells, G. L. (1980). What price justice? Exploring the relationship of lineup fairness to identification accuracy. *Law and Human Behavior*, *4*, 303–313.

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

Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology*, *66*, 482–489.

Metcalfe, J. (1996). Metacognitive processes. In E. L. Bjork & R. A. Bjork (Eds.), *Memory handbook of perception and cognition* (Vol. 10, pp. 381–407). San Diego, CA: Academic Press.

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

Murdock, B. B., & Dufty, P. O. (1972). Strength theory and recognition memory. *Journal of Experimental Psychology*, *94*, 284–290.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.

Smith, S. M., Lindsay, R. C. L., & Pryke, S. (2000). Postdictors of eyewitness errors: Can false identifications be diagnosed? *Journal of Applied Psychology*, *85*, 542–550.

Smith, S. M., Lindsay, R. C. L., Pryke, S., & Dysart, J. E. (2001). Postdictors of eyewitness errors: Can false identifications be diagnosed in the cross-race situation? *Psychology, Public Policy, and Law*, *7*, 153–169.

Sporer, S. L. (1992). Post-dicting eyewitness accuracy: Confidence, decision-times and person descriptions of choosers and nonchoosers. *European Journal of Social Psychology*, *22*, 157–180.

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology*, *78*, 22–33.

Sporer, S. L. (1994). Decision times and eyewitness identification accuracy in simultaneous and sequential lineups. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 300–327). New York: Cambridge University Press.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, *95*, 15–48.

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

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

Wells, G. L., Leippe, M. R., & Ostrom, T. M. (1979). Guidelines for empirically assessing the fairness of a lineup. *Law and Human Behavior*, *3*, 285–293.