Running Head: 10-12 SECOND RULE

Eyewitness Identification Accuracy and Response Latency: The Unruly 10-12 Second Rule

Nathan Weber

Neil Brewer

Flinders University

Gary L. Wells

Iowa State University

Carolyn Semmler

Amber Keast

Flinders University

Correspondence: Neil Brewer, School of Psychology, Flinders University, GPO Box 2100, Adelaide, South Australia 5001

Telephone: 61-8-8201 2725

Fax: 61-8-8201 3877

Email: neil.brewer@flinders.edu.au

Author Notes: This research was supported by grant A00104516 from the Australian Research Council to N. Brewer and G.L. Wells. Correspondence concerning this article should be addressed to Neil Brewer, School of Psychology, Flinders University, GPO Box 2100 Adelaide, South Australia 5001, Australia.

Abstract

Data are reported from 3,213 research eyewitnesses confirming that accurate eyewitness identifications from lineups are made faster than are inaccurate identifications. However, consistent with predictions from the recognition and search literatures, we did not find support for the "10-12 s rule" in which lineup identifications faster than 10-12 s maximally discriminate between accurate and inaccurate identifications (Dunning & Perretta, 2002). Instead, the time frame that proved most discriminating was highly variable across experiments, ranging from 5 s to 29 s, and the maximally discriminating time was often unimpressive in its ability to sort accurate from inaccurate identifications. We suggest several factors that are likely to moderate the 10-12 s rule.

Eyewitness Identification Accuracy and Response Latency: The Unruly 10-12 Second Rule

Since the mid 1990s, post-conviction use of forensic DNA evidence has lead researchers to believe that mistaken eyewitness identification is the primary cause of the conviction of innocent people (Wells et al., 1998). Years before forensic DNA was able to confirm this eyewitness misidentification problem, laboratory research by psychologists suggested that eyewitness identification evidence had two troublesome properties: (1) mistaken identifications can be quite high under some conditions and (2) once mistaken identifications occur, they are not easily distinguished from accurate identifications. The current article is concerned with the second of these problems: what is there to distinguish between accurate and inaccurate identifications? Specifically, it examines the “10-12 s rule” which proposes that lineup identifications faster than 10-12 s maximally discriminate between accurate and inaccurate identifications (Dunning & Perretta, 2002).

Research directed at finding “markers”, or assessment variables as they have been named by Sporer (1993), that might help crime investigators, judges, and juries distinguish accurate and mistaken eyewitness identifications has focused primarily on eyewitness confidence. Although early research, which relied on the point-biserial correlation to probe the confidence-accuracy relation (e.g., Bothwell, Deffenbacher, & Brigham, 1987; Wells & Murray, 1983), suggested that eyewitness identification confidence had little or no relation to eyewitness identification accuracy, research in recent years has been more promising. Using confidence-accuracy calibration, rather than the standard correlation statistic, a number of studies have reported impressive confidence-accuracy relationships using both eyewitness identification (Brewer, Keast, & Rishworth, 2002; Brewer & Wells, 2004; Juslin, Olsson, & Winman, 1996) and face recognition paradigms (Olsson, Juslin, & Winman, 1998; Weber & Brewer, 2003, 2004).

We are interested here in another promising marker of eyewitness identification accuracy that has emerged in recent years, namely identification response latency. Response latency[1] refers to the amount of time that an eyewitness takes to make and indicate their identification decision from a lineup. Research has shown that there is a negative statistical relation between response latency and accuracy in eyewitness identifications from lineups. Witnesses who make accurate identifications make their decision faster than witnesses who make inaccurate identifications (Dunning & Stern, 1994; Smith, Lindsay, & Pryke, 2000; Smith, Lindsay, Pryke, & Dysart, 2001; Sporer, 1992, 1993, 1994). Sporer (1992, 1993) suggested that lineup decisions are the result of a process of sequential comparison of each lineup member with an image of the offender in memory. When the target is compared with the image (i.e., a correct match), the large number of features in common allows a very fast decision. In contrast, a lineup foil (i.e., an innocent filler) will not have as many features in common with the image in memory and will, therefore, be matched more slowly. Thus, the negative relationship between accuracy and response latency will arise. This negative relation between response latency and identification accuracy appears to hold only for choosers; witnesses who correctly reject a target-absent lineup do not reach their decision faster than those who incorrectly reject a target-present lineup (Smith, Lindsay, & Pryke, 2000; Smith, Lindsay, Pryke, & Dysart, 2001; Sporer, 1992, 1993, 1994). Again, Sporer argued that this can be explained by the sequential matching decision process. According to his argument, as all rejections–

regardless of accuracy–require all lineup members to be considered, rejections should be uniformly slow. This is similar to the pattern shown with confidence in that the relation to accuracy is primarily manifested among choosers (sorting hits from false alarms) rather than non-choosers (sorting correct rejections from misses).

The potential utility of this negative relation between latency and accuracy is considerable for two reasons. First, unlike confidence, response latency is a performance variable rather than a self-report variable. As a self-report variable, confidence is subject to a variety of distortions (Semmler, Brewer, & Wells, in press; Wells & Bradfield, 1998). Response latency, on the other hand, is a natural product of performing the lineup identification task and response latency can be measured even without the witness's awareness that it is being measured. Second, although confidence and decision latency are (negatively) correlated, they are not fully redundant in accounting for variance in accuracy (Smith et al., 2001). Hence, together, confidence and latency could serve as better markers of eyewitness identification accuracy than either alone.

Importantly, knowledge of the negative accuracy-decision latency relationship, although potentially informative about the cognitive processes involved in identification decisions, is of little practical use when assessing the likely accuracy of any single identification decision. For example, if a witness makes an identification in 15 s, how are the police, judges, or jurors to know if this is a fast (and, therefore, likely to be accurate) or a slow (and, therefore, likely to be inaccurate) decision? One possible solution to this problem is the identification of the discriminant function for classification of accurate and inaccurate identifications (Smith et al., 2000, 2001; Sporer, 1994), but this approach is limited if a different function must be derived for each stimulus array. Dunning and

Perretta (2002) identified an alternative solution. They suggested a new analytic strategy, the time-boundary analysis, that could identify the time boundary that best distinguishes accurate from inaccurate identifications. The identification of such a time boundary, if it is stable, is of obvious practical import because it establishes a simple criterion by which to evaluate the identification.

The time-boundary analysis is based on the examination of the difference in the proportion of correct identifications made before a specific time and the proportion of correct identifications made after that time. This comparison can be made by computing the chi-square statistic for the appropriate 2 (time boundary: faster than boundary vs. slower than boundary) × 2 (accuracy: correct vs. incorrect) contingency table. By, systematically examining the chi-square value produced at each of a series of values (e.g., 1 s, 2 s, 3 s, and so on), the time boundary that best distinguishes accurate from inaccurate identifications can be identified as the boundary that produces the greatest chi-square value. The analysis can also be conducted by computing the log odds ratio instead of a chi-square value.

Dunning & Perretta (2002) performed time-boundary analyses on each of four data sets: two of these (Ns = 41 and 50) derived from experiments designed for time boundary analyses, the other two (Ns = 221 and 96) were from experiments designed for other purposes. Based on these analyses they concluded that a 10-12 s time boundary was most useful at distinguishing positive identifications that were highly likely to be accurate from positive identifications that were less likely to be accurate. Importantly, although witnesses who made a positive identification before the 10-12 s time boundary had a probability of being correct approaching 90%, those who responded after 12 s were accurate in approximately 50% of cases. There is no doubt that the data that Dunning and Perretta (2000) analyzed show an impressive ability of the 10-12 second time frame to discriminate between accurate and inaccurate identifications. In fact, Dunning and Perretta referred to this as the “10-12 second rule,” an indication of their confidence that the 10-12 s time frame would prove stable across conditions of various types.

The idea of a stable time boundary is consistent with Dunning and Stern’s (1994) notion of automatic and deliberative processing in eyewitness identification decision making. They argue that automatic (i.e., fast and unconscious) decision processes are likely to have been characteristic of witnesses who make accurate positive identifications. In contrast, more conscious, deliberative, time consuming decision processes are likely to be used by witnesses who make inaccurate positive identifications. Because automatic judgments are typically insensitive to changes in the decision context, Dunning and Perretta (2002) argue that the response latency of automatic decisions is likely to be constant across lineups despite changes in the latency of deliberative or process-of-elimination decisions. Thus, a concentration of automatic, and consequently predominantly accurate, decisions will consistently be observed at low latencies, that is, before the 10-12 s time boundary. It is important to note that the automatic processing described by Dunning and Stern clearly differs from the rapid and uncontrolled automatic processing, that does not necessitate attention, described by Shiffrin and Schneider (1977; Schneider & Shiffrin, 1977). Obviously, the time scale of eyewitness identification decisions is typically much longer than the response latencies observed in the search and detection experiments of Shiffrin and Schneider. Although this time scale is not compatible with their classification ‘automatic’, we certainly do not wish to suggest that some identifications cannot occur with much less apparent effort or conscious awareness of the underlying decision process than others. Regardless, not only does Dunning and Perretta's 10-12 s rule raise an important practical issue, but it also highlights an important theoretical question. Specifically, does the cognitive process underlying identification decisions produce fast and accurate identifications with response latency that is not influenced by the nature of the stimulus or test conditions?

Existing theoretical and empirical work in the areas of memory and visual search – admittedly not involving eyewitness identification tests – suggest to us that the answer to this question is almost certainly no. Consider, for example, a model of search processes such as that outlined by Treisman and colleagues (Treisman & Gelade, 1980; Treisman & Gormican, 1988). Here, parallel processing or search of the entire stimulus array may lead to an early preattentive or automatic registration of some key feature – or indeed, in the case of a face, a featural configuration – that defines the target's location and leads to 'popout' of the target stimulus. In the lineup situation this could indeed produce a very rapid or automatic identification, especially if the feature was something very obvious like a scar or tattoo that clearly did not appear on the lineup foils. But this could also occur for more complex configurations of features that are not readily amenable to description, perhaps again not captured in lineup foils, but provide an important basis for the witness's discrimination. The response may also be accurate, although given that faces are complex stimuli, there would appear to be ample scope for error, even if it was a holistic image of a face that apparently popped out. Should parallel processing not produce such an outcome, however, a serial and potentially more time consuming search of the display would commence. If this serial search located an extremely strong match for the image held in memory, the search may well be terminated. The outcome would likely be an accurate but this time somewhat slower identification, with the actual latency depending on the arrangement of the lineup, its size, the position of the target, and so on. (Note that this identification could be very fast if the match was located at the beginning of the search.) But what if the image held in memory is poor quality or the lineup contains many closely competing foils or plausible matches (or both)? Then an exhaustive search (or multiple searches) of the whole array is likely, a search that would be relatively slow and could produce a correct or an incorrect response. (Note that different search models might lead to variations on the above idea.s)

What might all of this mean for identification latency and the optimum time boundary? Clearly identification latency would be expected to vary across individuals and situations, depending on a wide range of factors such as the quality of the image held in memory (at least in part reflecting encoding conditions), the degree of match between the image and the lineup stimulus, the composition of the lineup with respect to similarity of foils, the number of lineup members, the visual angle subtended by the lineup, the witness's decision criterion, etc. None of this, of course, precludes the possibility of a parallel or automatic detection process that is both fast and accurate. But nor does it preclude the possibility of fast, incorrect identifications, slower but accurate identifications, or the possibility that characteristics of the stimulus, the image, and the lineup will almost certainly shape the distribution of response latencies for accurate and inaccurate identifications. Thus, we would also expect the time boundary – and the proportion of correct responses within that boundary – to vary.

Similarly problematic for the 10-12 s rule are cognitive theories and empirical data from a number of domains outside of eyewitness identification that suggest that a constant time boundary is unlikely. For example, the information accumulation model of recognition memory outlined by Van Zandt (2000) is described in enough detail to make very specific predictions regarding response latency and accuracy: specifically, when people can regulate their own response times (i.e., they are not instructed to respond within a certain time), decreases in discriminability (i.e., changes that make it harder to discriminate old from new stimuli) will increase decision latency. Importantly, this increase is predicted for both accurate and inaccurate decisions. Thus, accurate recognition decisions will still be made, on average, faster than inaccurate decisions, but the speed of both will decrease with decreasing discriminability. Therefore, Van Zandt's model would predict that the time boundary for identification decisions should be sensitive to changes in stimuli, viewing conditions, and testing conditions that influence the ease with which witnesses can discriminate the offender from foils or innocent suspects. Such a prediction is also supported by empirical observations in other task domains. For example, psychophysicists have consistently shown variations in correct and error response times with various experimental manipulations (e.g., Festinger, 1943; Vickers & Packer, 1982).

Indeed at first glance, consideration of recent response latency data from eyewitness identification studies appears to support our position and challenge the likely generality of the 10-12 s rule. For example, Sporer (1993) found that correct identifications were made with a mean response latency of 3.61 s and incorrect positive identifications with a mean response latency of 8.06 s, whereas in Sporer (1992) the response latencies for correct and incorrect positive identifications were 10.69 s and 23.76 s, respectively. Thus, the latency data appear to suggest that (a) response latencies of both correct and incorrect identifications are variable and (b) incorrect identifications can be made within the 10 s time boundary. Importantly, neither of these findings are necessarily incompatible with the 10-12 s rule or the notion of automatic versus deliberative processing. For example, these fluctuations in mean response latency for correct and incorrect decisions could be due solely to variations in the mean response latency for deliberative decisions. Or in other words, accurate and consistently fast automatic decisions could be present in any of these data sets. Furthermore, the 10-12 s time boundary could optimally discriminate correct from incorrect identifications for any combination of mean response latencies in which correct identifications are made, on average, faster than incorrect identifications. In other words, without knowledge of the distribution of response latencies for correct and incorrect identifications, no conclusion about the optimal time boundary can be drawn from means and standard deviations alone. Thus, although the 10-12 s rule may, at first, appear to be obviously wrong, this conclusion can not be supported without conducting a time boundary (or equivalent) analysis.

In light of these findings and predictions, a test of the generality of Dunning and Perretta’s 10-12 s rule is necessary for two reasons. First, given the potential practical utility of the 10-12 s rule and the appeal that such a simple and apparently powerful tool could have amongst law enforcement, a thorough understanding of the breadth of its applicability is warranted. Second, such a test provides an indication of the more appropriate theoretical model on which to base theorizing about identification decisions. Specifically, support for the 10-12 s rule would suggest that Dunning and Stern’s account of highly accurate automatic identifications with invariant response latency is fruitful. In contrast, a negative finding for the 10-12 s rule suggests that theories predicting that response latency for both correct and incorrect decisions vary with testing conditions are likely to form a more useful theoretical foundation for future eyewitness identification work. To accomplish this we have reanalyzed data collected from four different experiments in our own laboratory, using a different set of stimulus events and lineups to those used by Dunning and Perretta. Importantly, these data sets are all very large and therefore should provide a good indication of the true optimum time boundary for the given experimental conditions and stimuli. These data sets have a number of other valuable features. First, they provide several replications with the same stimulus event and lineups; a failure to detect a stable time boundary across these replications would present a problem for the 10-12 s rule. Second, the experiments employed different retention intervals and included two different target stimuli (and consequently two different lineups) and thus provide a good test of the 10-12 s rule across different witnessing and testing conditions. Third, children participated as witnesses in one of the four experiments, providing a test of the generality of the 10-12 s rule across different witness ages.

In addition to investigating the location of the time boundary itself, our other major interest was in the accuracy rates for responses faster versus slower than the identified boundary. A replication of Dunning and Perretta's (2002) rate of almost 90% accurate identifications among witnesses who respond faster than the time boundary would provide very strong support for the practical utility of the time boundary as diagnostic of accurate positive identifications. In contrast, the identification of lower accuracy rates before the time boundary would limit the extent to which the time boundary could be accepted as practically useful. For example, if only 60% of identifications made before the boundary were accurate, categorizing these fast identifications as likely to be accurate would not appear prudent. Thus, the other major focus of the paper was to explore the proportion of accurate identifications made within Dunning and Perretta's 10-12 s time boundary and (if a different boundary was identified) within the time boundary identified in these data.

Method

We drew upon data that were collected primarily for the purpose of studying eyewitness confidence. Most of these experiments were designed to address the issue of confidence-accuracy calibration, a procedure which requires unusually large sample sizes. Even though the calibration question did not require the measurement of response latency, response latency was nevertheless measured in each of these experiments. Accordingly, we were able to analyse response latency data from 3,213 experimental witnesses across four experiments for purposes of the current article. None of the previous articles or manuscripts relating to these four experiments report any analyses of the response latency data.

The crime video depicted the theft of a credit card from the front counter of a restaurant. The video lasted for 140 s. Two people were featured in the video, the thief and the waiter from whom the credit card was stolen. Both were shown from various angles, with a full or partial view of the thief's face available for 23 s and of the waiter's for 72 s. Some of the studies (i.e., Studies 1 and 2) asked participants to identify the thief only, while the remaining studies (i.e., Studies 3 & 4) required participants to identify both the thief and the waiter. A detailed description of the stimulus video can be found in Brewer et al. (2002). In sum, these studies provide four replications with the same stimulus materials while also providing data for two different target stimuli and lineups.

In all studies and for both targets the identification task was computerized. Participants were presented with a 2 × 4 array of photographs. They were displayed on a 15-in. monitor with the resolution set at 1,024 × 768 pixels. The photographs were presented in colour and displayed with an onscreen size of 4 × 5.57 cm.

Study 1

Overview. Study 1 is a reanalysis of data collected for Brewer et al. (2002) investigation of confidence-accuracy calibration in eyewitness identification. Specifically, they examined the impact on confidence-accuracy calibration of two interventions that were designed to improve calibration. Because the interventions were not implemented until after the participants had made their identification decisions, the intervention manipulations could not have influenced the latency-accuracy relationship and, thus, are not detailed here.

Participants. The 944 participants were recruited from an undergraduate volunteer register, attendees at a university information day, and from the student employment service. Participants recruited from the employment service were paid for their participation.

Procedure. Before beginning the experiment participants were informed that they would be participating in a forensic psychology study. They were then shown the stimulus video in small groups (i.e., 2-4). After viewing the video they worked individually on a filler task for 20 min before completing the computerized identification task. Participants were explicitly instructed that the thief may or may not be present in the lineup they were about to be shown. A target-present or target-absent eight member lineup was then presented to the participant who was instructed to click on the button under the face they thought was the thief or on the 'Not Present' button if they thought the thief was not in the lineup. The response latency (i.e., the elapsed time from lineup exposure to the click on a response button) was recorded by the computer. Participants then completed another 5 min pen-and-paper task, which varied depending on the experimental condition to which they were assigned, after which they rated their confidence in the accuracy of their decision on an 11-point, 0% - 100% scale.

Study 2

Overview. Study 2 is a reanalysis of data reported by Semmler et al. (in press). They investigated the impact of confirming feedback on participants' confidence in the accuracy of their identification decision. Again, this manipulation occurred after the identification decision, so the manipulation could not have influenced the relation between identification accuracy and identification latency.

Participants. Four hundred and sixteen participants were recruited from the student employment service and paid for their participation.

Procedure. This procedure departed from that used in Study 1 in only two important ways. First, after viewing the stimulus video participants worked on the filler task for 15, not 20, min. Second, after making an identification decision, and hence after the recording of latency, participants were given confirming or no feedback and then asked to rate both their current and retrospective confidence in the accuracy of their decision.

Study 3

Overview. Study 3 is a reanalysis of data collected by Brewer and Wells (2004) in another investigation of variables affecting confidence-accuracy calibration.

Participants. Twelve hundred participants completed this study. They were recruited from undergraduate and community groups and all were paid for their participation.

Procedure. The basic procedure used in Study 3 differed from Study 1 in three ways. As in Study 2, the retention interval, during which a filler task was completed, lasted only 15 min. Second, participants rated their confidence immediately after making an identification decision. Third, after making a decision, and rating their confidence, about the thief lineup, participants were asked to identify the waiter from a lineup and, subsequently, to rate their confidence in the accuracy of their decision. As with the thief, eight-person target-present and target-absent lineups were used for the waiter. In addition to these differences, two experimental manipulations were used in a between-subjects design. First, before viewing the lineup participants were given biased (i.e., failure to warn about possibility of target being absent) or unbiased (as in Studies 1 and 2) lineup instructions. Second, two selections of lineup foils who were either high or low in similarity to the target were created for each stimulus (i.e., the thief and the waiter) and presented with the target in target-present lineups or the target’s replacement in target-absent lineups.

Study 4

Overview. Study 4 is a reanalysis of data collected by Keast, Brewer, and Wells (2004, Experiment 1) in their investigation of children’s eyewitness identification performance and the confidence-accuracy relation.

Participants. Six hundred and fifty three primary and middle school children (age *M* = 11y 8m, *SD* = 10.0m) completed the study.

Procedure. The procedure used in Study 4 differed from that employed in Study 3 in two ways. Participants in this study viewed the video in class groups, ranging from 9 to 26 children, and were given high similarity lineups.

Results

The distributions of identification responses for the four studies are displayed in Table 1. Table 2 displays the descriptive statistics for untransformed response latency and the inferential statistics (and associated effect size measure, Cohen's d) for comparisons of transformed latency scores (square root for study 1, and logarithmic for studies 2, 3, and 4). Consistent with previous research, correct positive identifications were found to be made significantly faster than incorrect positive identifications, with effect sizes ranging from small to moderate. For all samples the response latency for correct non-choosers did not differ significantly from that of incorrect non-choosers, thus non-choosers data are not considered further.

Time-boundary analyses were conducted for all samples to establish the maximally discriminating time for separating accurate from inaccurate witnesses. Consistent with the technique used by Dunning and Perretta (2002), we computed a chi-square statistic based on the 2 (time boundary: faster or equal vs. slower) × 2 (accuracy: correct vs. incorrect) contingency tables with the time boundary set at each integer value from 1 s to 40 s (i.e., 1 s, 2 s, 3 s, and so on until 40 s). Non-choosers were not included in these analyses, that is, only positive identification decisions were examined. Following Dunning and Perretta (2002), the peak in this series of chi-square values should indicate the time boundary that best discriminates correct from incorrect responses. An important limitation of the time-boundary analysis is the subjective nature of the identification of peaks in the chi-square by time-boundary curves. Following Dunning and Perretta we identified the peaks as occurring at the time-boundary with the greatest chi-square value for unimodal curves and the time-boundaries with the greatest chi-square values for each mode in multimodal curves. An important step, if the time-boundary analysis is to become an important tool for researchers, is the development of an objective method for identification of these peak time-boundaries. One approach to this is to report a confidence range about each chi-square peak, defined as the range of time boundaries with chi-square values within ± 1 standard error of the peak value.

Plots of chi-square value (and standard error) by time-boundary for all samples are displayed in Figure 1. Following is a list of the peaks identified for each sample, the confidence range and the phi coefficient, the square of which can be interpreted as eta-square, as an indication of the effect size for the difference in proportions correct at either side of the peak time-boundary. The one striking feature of these data is that, unlike Dunning and Perretta, no single time window can be associated with the chi-square peaks for all samples. Study 1 shows a peak at 24 s (confidence range = 23-25 s, $\phi = .19$) and also at 28 s (confidence range = 28-29 s, $\phi = .19$), and study 2 shows peaks at 12 s (confidence range = 12 s, $\phi = .21$) and 17 s (confidence range = 17 s, $\phi = .21$). For the study 3 data, peaks are evident at 14 s (confidence range = 13-14 s, $\phi = .26$) for the thief lineup and 10 s (confidence range = 8-11 s, $\phi = .28$) for the waiter. For study 4 peaks are evident at 16 s (confidence range = 16-17, $\phi = .12$) and 27 s (confidence range = 21-28 s, $\phi = .13$) for the thief lineup and 5 s (confidence range = 5 s, $\phi = .25$) for the waiter lineup.

It is also noteworthy that this lack of consistency is not merely due to erratic performance of the child witnesses as no single peak can be identified when only the adult samples (i.e., Studies 1, 2, and 3) are considered. In regards to the specific 10-12 s time boundary identified by Dunning and Perretta (2002), only two of the six samples examined revealed a chi-square peak that was within the 10-12 s window, or a chi-square range that included the 10-12 s window. Thus, these data provide an unequivocal demonstration that the optimum time-boundary for discriminating correct from incorrect positive identifications is not constant across all conditions and participants.

Our second major focus was addressed by two further analyses to identify the proportions correct before and after (a) the 10 s time boundary identified as optimum by Dunning and Perretta (2002), and (b) the time boundary as suggested by the earliest peak chi-square value observed in Figure 1 (i.e., the empirical time boundary). These accuracy rates are displayed in Table 3. The most important feature of these accuracy rates is that in none of the six data sets did the accuracy rate for identifications made inside the time boundary approach the high (approximately 90% across studies) accuracy observed by Dunning and Perretta. In fact, fewer than six out of every ten identifications within the time boundary were accurate. This is true whether Dunning and Perretta’s 10 s time boundary (% correct = 51.2) or the boundary associated with the chi-square peak is used (% correct = 57.8). Even after excluding the child samples the overall accuracy rates before the boundary are only 65.5% (10 s boundary) and 66.9% (empirical time boundary). Thus, these data suggest that, contrary to Dunning and Perretta’s findings, even when the time boundary that best discriminates accurate from inaccurate responses is used, this technique does not always identify responses that are likely to be accurate.

Does confidence help to further discriminate between accurate and inaccurate identifications within the optimal time boundary, as Smith et al.'s (2001) and Sporer's (1994) findings suggest? Owing to the poor relationship between confidence and accuracy displayed by the child sample (Keast et al., 2004) we did not conduct such an analysis for the child sample. Further, data from study 2 (Semmler et al., in press) participants who received post identification feedback were not included in this analysis because their confidence was artificially manipulated in that study. For the remaining studies, confidence estimates of 90% or 100% on the 0-100% scale were categorized as highly confident for the purpose of this analysis. Table 4 contrasts the accuracy rates for identifications made with both high confidence and within the (10 s or empirical) time boundary with the remaining responses made within the time boundary but with lower confidence. Although few fast and highly confident responses were made in some samples, the accuracy rates are substantially higher than those obtained when only the time boundary is used to categorize responses. This is particularly evident when the overall accuracy rates (not considering confidence) for fast responses (10 s: 65.5%; empirical: 66.9%) and for fast less confident responses (10 s: 53.5%; empirical: 62.5%) are compared with the overall accuracy for fast and highly confident responses (10 s: 88.1%; empirical: 84.3%).

Discussion

The purpose of this study was to test the generality of Dunning and Perretta's (2002) two major findings that (a) a 10-12 s time boundary best distinguishes accurate from inaccurate positive identification responses and (b) the positive responses made before this boundary were likely (almost 90%) to be correct and, thus, provide a test of the notion of accurate automatic responders with invariant response latency. The data from the adult and child samples analyzed here demonstrate that the time boundary and accuracy rates are not invariant for (a) replications with the same target stimuli and lineup structure, (b) different event stimuli and lineup structures, and (c) age ranges encompassing children and adults. Further, and just as important, regardless of whether the time boundary identified by Dunning and Perretta (i.e., 10 s) or our empirically determined optimum time boundaries were used to classify identification responses, the proportion of accurate fast identifications was relatively poor, with only one of the eight samples exceeding 80%. For the 10-12 s rule to be a useful tool for law enforcement the proportion of accurate identifications inside the time boundary is critical. Thus, as the 10-12 s rule does not reliably diagnose identifications with a high probability of accuracy, its adoption in the legal setting is not justified.

In addition to their practical implications, these findings have important ramifications for our understanding of identification decision processes. Dunning and Perretta (2002) argued that the consistent optimal time boundary was a function of correct identifications largely being the result of an automatic process, the latency of which is unaffected by the nature of the event and the lineup. However, this view is not consistent with either the optimal time-boundaries identified for these data or with the average response latencies for accurate and inccurate identifications. Figure 2 clearly shows that the conditions that produced changes in identification latency produced parallel changes in both correct and incorrect identification latencies and in the location of the optimum time boundary. These results are not surprising given previous research on automatic and controlled search processes. Furthermore, they are consistent with Van Zandt's (2000) information accumulation model of recognition memory which suggests that the location of the optimum time boundary will vary with the discriminability, or difficulty, of the identification task. Thus, a successful theory of the identification decision process appears likely to come from models of recognition memory, like Van Zandt's, that provide some basis for predicting variations in correct and incorrect response latencies with variations in encoding and retrieval conditions.

Our data sets failed to show that witnesses who respond within the 10 s time boundary (or within the empirically established time boundary) have impressively high accuracy rates. However, a striking feature of our data is that impressive accuracy rates were obtained when high confidence and the 10 s time boundary were used together as a marker of accuracy for the adult samples. Specifically, the combined use of 90-100% confidence and the 10 s time boundary diagnosed identification decisions with a high probability of accuracy (88.1% overall). Importantly, this discrimination was superior to that achieved by the use of either the time boundary (i.e., empirical optimum or 10 s) or confidence alone. This finding is consistent with the results of Smith et al. (2000, 2001) and Sporer (1994) who demonstrated with discriminant functions analyses that latency and confidence, along with other factors, combined to significantly discriminate accurate from inaccurate identifications. Like Dunning and Perretta's (2002) 10-12 s rule, support for the predictive utility of the combination of high confidence and a rapid identification may not be found across all combinations of stimuli and witnessing conditions. However, it does serve to underscore the importance of developing a theoretical understanding of the decision process and the potential utility of pursuing combined markers of identification accuracy.

An important aspect of an understanding of the decision process that is highlighted by this research is the identification of factors that influence the optimum time-boundary or the response latency of correct and incorrect identifications. Many potentially influential factors exist. Earlier we talked about factors such as the quality of the image held in memory (something that will vary with the encoding conditions experienced by different witnesses), and the witness's decision criterion (likely to be influenced by lineup expectations and instructions). Another obvious factor is the type of lineup procedure employed. Our data were obtained from simultaneous lineups where participants were able to view all lineup members and compare them with each other before making a decision. The restriction on this comparison imposed by the sequential lineup procedure (Lindsay & Wells, 1985), and the requirement for a decision to be made about each lineup member presented alone, could drastically change the type of processing required of the participant and also the relationship between response latency and accuracy. Further, factors such as live versus computerized or photo array presentation of the lineup to participants could impact response latency, and subsequently the optimum time-boundary, as participants may take longer to process the greater detail present in a live lineup. Perhaps the most important group of potentially influential factors, however, are those that result from the nature of the offender and the characteristics of the lineup (i.e., in the case of laboratory experiments, from the stimulus materials). A theory that describes the impact of lineup structure or characteristics of the suspect's appearance on the relationship between response latency and accuracy could have enormous practical benefits for the utility of response latency as a marker of identification accuracy. Therefore, understanding the impact of factors such as these on response latency and, more importantly, on the relationship between response latency and identification accuracy is an important goal for future work.

In sum, these data provide clear evidence that a 10-12 s time boundary does not always optimally discriminate accurate from inaccurate identifications and, more importantly, that identifications made within 10 s are not always highly likely to be accurate. Consequently, the practical adoption of a 10-12 s rule for identifying correct identification decisions is unjustified. However, the potential utility of the combination of latency and confidence as markers of accuracy has been demonstrated. Finally, these data are not consistent with Dunning and Stern’s notion of automatic and deliberative processing and we suggest that a more fruitful foundation for theories of the identification decision process is likely to come from recognition memory models that predict an impact of witnessing and lineup conditions on the response latency of correct and incorrect positive identifications.

References

Bothwell, R. K., Deffenbacher, K. A., & Brigham, J. C. (1987). Correlations of eyewitness identification accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology*, *72*, 691-695.

Brewer, N., Keast, A., & Rishworth, A. (2002) Improving the confidence-accuracy relation in eyewitness identification: Evidence from correlation and calibration. *Journal of Experimental Psychology: Applied, 8*, 44-56.

Brewer, N., & Wells, G. L. (2004). *The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity and target-absent base rates*. Manuscript in preparation.

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 identifications via inquiries about decision processes. *Journal of Personality and Social Psychology, 67*, 818-835.

Festinger, L. (1943). Studies in decision: I. Decision time, relative frequency of judgement, and subjective confidence as related to physical stimulus difference. *Journal of Experimental Psychology*, *32*, 291-306.

Juslin, P., Olson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can and cannot be inferred from a low confidence-accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 5*, 1304-1316.

Keast, A., Brewer, N., & Wells, G. L. (2004). *Children's eyewitness identification performance and the confidence-accuracy relation*. Manuscript in preparation.

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.

Olsson, N., Juslin, P., & Winman, A. (1998). Realism of confidence in earwitness versus eyewitness identification. *Journal of Experimental Psychology: Applied*, *4*, 101-118.

Schneider, W., & Shiffrin, R.M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*, 1-66.

Semmler, C., Brewer, N., & Wells, G. L. (in press). Effects of postidentification feedback on eyewitness identification and nonidentification confidence. *Journal of Applied Psychology*.

Shiffrin, R.M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*, 127-190.

Smith, S. M., Lindsay, R. C. L., & Pryke, S. (2000). Postdictors of eyewitness errors: Can false identification 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 non-choosers. *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, and M. P. Toglia (Eds.) pp. 300-327. *Adult eyewitness testimony: Current trends and developments*. New York: Cambridge University Press.

Treisman, A. M., & 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.

Vickers, D., & Packer, J. S. (1982). Effects of alternating sets for speed and accuracy on response time, accuracy, and confidence in a unidimensional discrimination task. *Acta Psychologica*, *50*, 179-197.

Weber, N., & Brewer, N. (2003). The effect of judgment type and confidence scale on confidence-accuracy calibration in face recognition. *Journal of Applied Psychology*, *88*, 490-499.

Weber, N., & Brewer, N. (2004). *Confidence-accuracy calibration in absolute and relative face recognition judgments*. Manuscript submitted for publication.

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.

Wells, G. L., & Murray, D. M. (1984). Eyewitness confidence. In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony: Psychological perspectives* (pp. 155-170). New York: Cambridge University Press.

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.

Footnote

[1] We follow convention from the processing speed literature in referring to the time that elapses from presentation of the lineup to when the participant indicates their choice as “response latency,” rather than decision time or latency, because it incorporates both decisional and motor components.

Table 1

Identification Response Frequencies (and Percentages) for Target Present and Target Absent Lineups

	Target Present			Target Absent	
Study: Target	Correct ID	Incorrect ID	Incorrect Rejection	Incorrect ID	Correct Rejection
1: Thief	292	121	392	38	101
	(36.3%)	(15.0%)	(48.7%)	(27.3%)	(72.7%)
2: Thief	70	33	105	46	162
	(33.7%)	(15.9%)	(50.5%)	(22.1%)	(77.9%)
3: Thief	222	105	274	197	402
	(36.9%)	(17.5%)	(45.6%)	(32.9%)	(67.1%)
3: Waiter	367	132	100	329	272
	(61.3%)	(22.0%)	(16.7%)	(54.7%)	(45.3%)
4: Thief	66	177	101	206	103
	(19.2%)	(51.5%)	(29.4%)	(66.7%)	(33.3%)
4: Waiter	154	103	52	276	68
	(49.8%)	(33.3%)	(16.8%)	(80.2%)	(19.8%)

Table 2

Means and Standard Deviations for Response Latency (in s) for Accurate and Inaccurate Positive Identifications

Study: Target	Statistic	Identification Accuracy: Accurate	Identification Accuracy: Inaccurate	Statistic: *t*	Statistic: *df*	Statistic: *d*
1: Thief	*M*	26.39	33.77	-4.12*	266.19	0.38
	SD	14.69	20.62			
2: Thief	*M*	24.29	27.45	-2.15*	147	0.36
	SD	22.68	16.94			
3: Thief	*M*	18.54	25.99	-6.77*	522	0.60
	SD	11.93	17.11			
3: Waiter	*M*	13.83	20.24	-8.97*	826	0.62
	SD	11.03	13.71			
4: Thief	*M*	20.71	26.80	-2.92*	126.33	0.34
	SD	14.27	22.12			
4: Waiter	*M*	11.34	15.25	-5.39*	554	0.50
	SD	10.26	11.59			

*p < .05

Table 3

Accuracy Rates for Identifications Before and After the 10 s and Empirically Determined Time Boundaries

	10 s Time Boundary		Empirical Time Boundary	
Study: Target	% Correct Before	% Correct After	% Correct Before	% Correct After
1: Thief	78.6	64.3	74.2	56.0
2: Thief	61.1	45.0	67.7	41.5
3: Thief	69.3	37.9	61.7	33.7
3: Waiter	64.1	34.6	64.1	34.6
4: Thief	22.4	15.3	20.2	11.4
4: Waiter	37.4	21.7	58.8	24.5
Overall	51.2	36.4	57.8	31.8

Table 4

Accuracy Rates for High Confidence and Lower Confidence Identifications Before and After the 10 s and Empirically Determined Time Boundaries

		10 s Time Boundary		Empirical Time Boundary	
Study: Target	Confidence – Time boundary	% Correct	*N*	% Correct	*N*
Overall	High – Before	88.1	135	84.3	204
	High – After	69.0	155	61.8	89
	Low – Before	53.5	243	62.5	502
	Low – After	60.0	949	37.0	1116
Study 1: Thief	High – Before	100.0	7	80.4	51
	High – After	78.7	61	82.4	17
	Low – Before	57.1	7	72.3	166
	Low – After	62.0	376	66.8	217
Study 2[a]: Thief	High – Before	100.0	3	100.0	5
	High – After	80.0	5	50.0	6
	Low – Before	33.3	6	50.0	8
	Low – After	44.9	49	40.9	44

Table 4 (cont.)

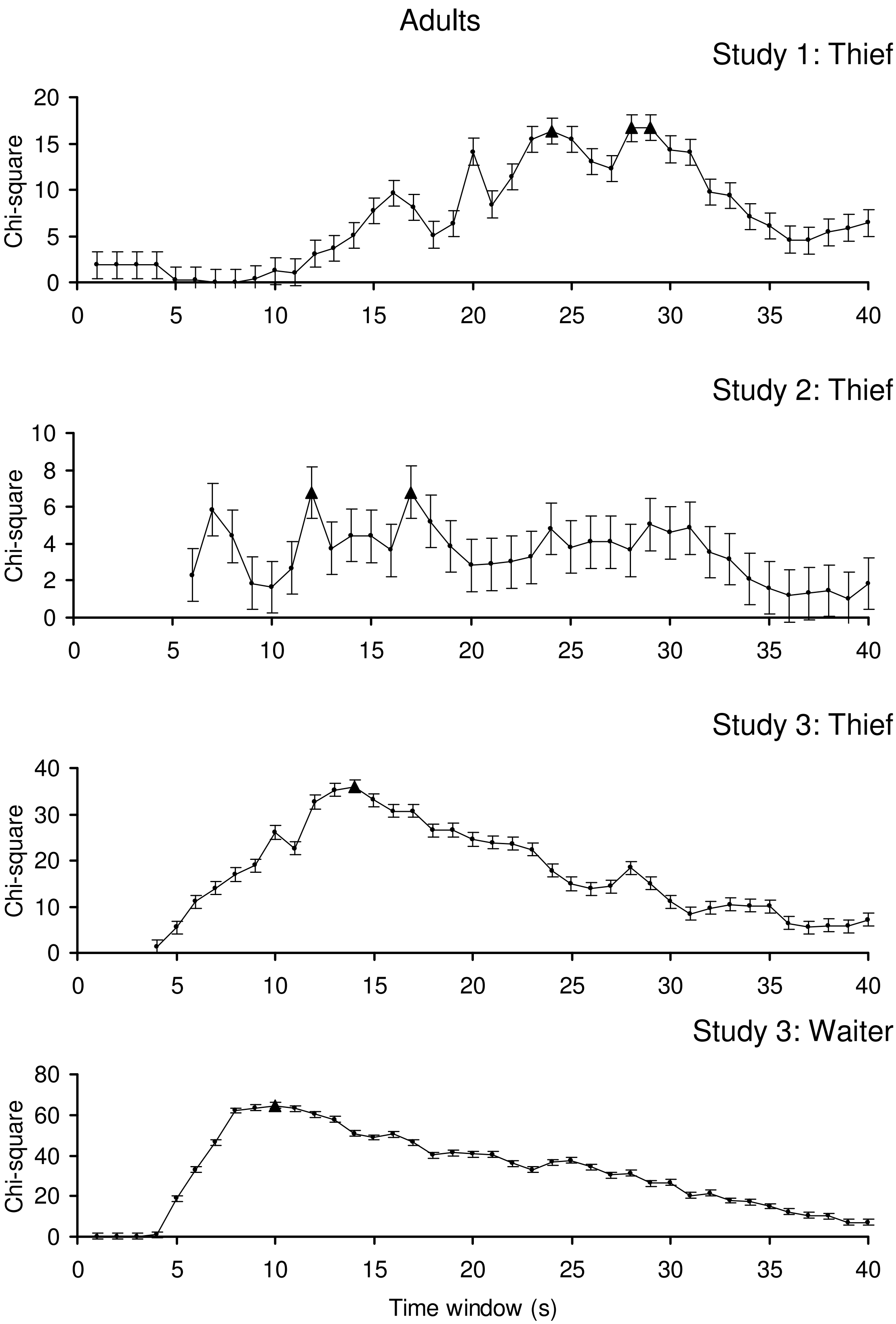
Study 3: Thief	High – Before	86.2	29	78.7	47
	High – After	64.4	45	63.0	27
	Low – Before	58.7	46	54.8	115
	Low – After	34.9	404	31.3	335
Study 3: Waiter	High – Before	86.8	91	86.8	91
	High – After	54.3	35	54.3	35
	Low – Before	52.7	182	52.7	182
	Low – After	33.3	520	33.3	520

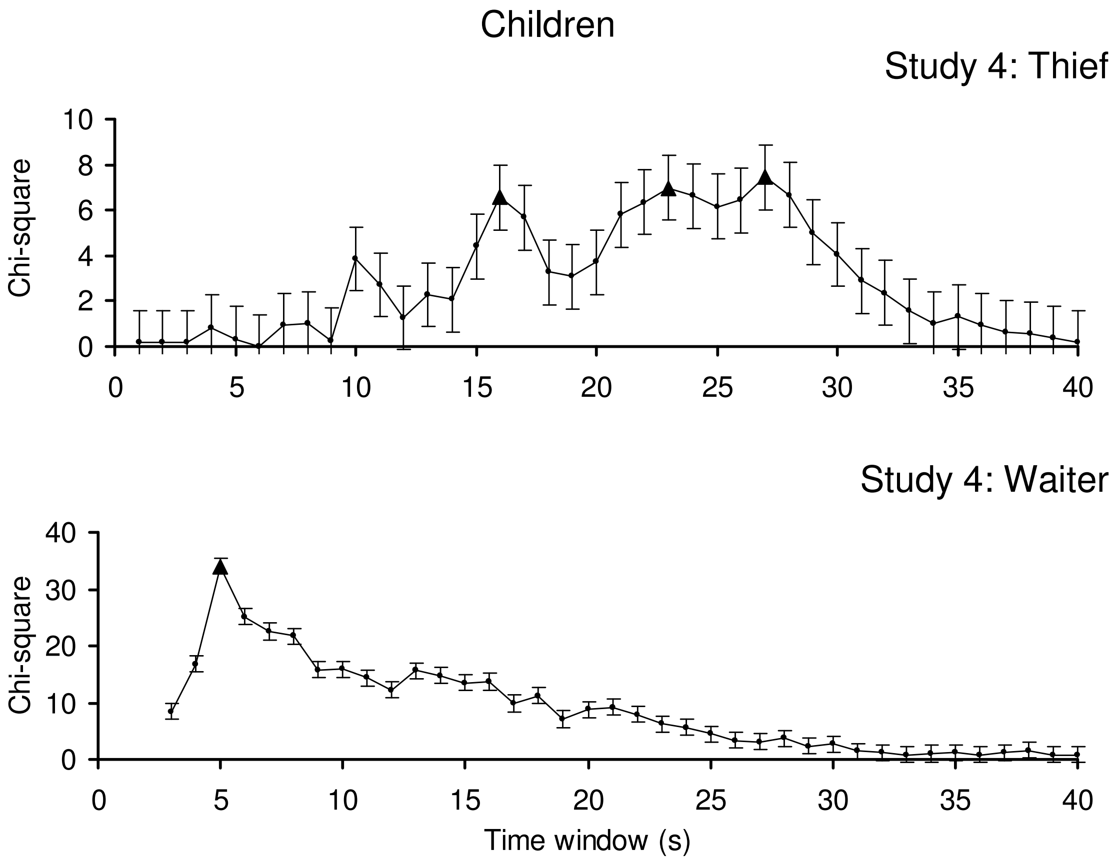
[a] Participants in this study provided two confidence estimates (current and retrospective) in counterbalanced order. Only the first reported measure was used in this analysis.

Figure Captions

Figure 1. Plots of chi-square (and standard error) by time-boundary for each study. Note that to allow easy discrimination of the Chi-Square peak or peaks for each curve the y-axis scales are not consistent.

Figure 2. Plots of correct and incorrect latency (and standard error) and optimum time boundary (and confidence range) for thief and waiter targets in Studies 1-4. Note that the confidence range sometimes does not extend beyond the observed value, hence the apparent absence of some error bars.


Adults
Study 1: Thief
Chi-square
0
5
10
15
20
Study 2: Thief
Chi-square
0
2
4
6
8
10
Study 3: Thief
Chi-square
0
10
20
30
40
Study 3: Waiter
Chi-square
0
20
40
60
80
0
5
10
15
20
25
30
35
40
Time window (s)



Children

Study 4: Thief

Study 4: Waiter

Figure 1

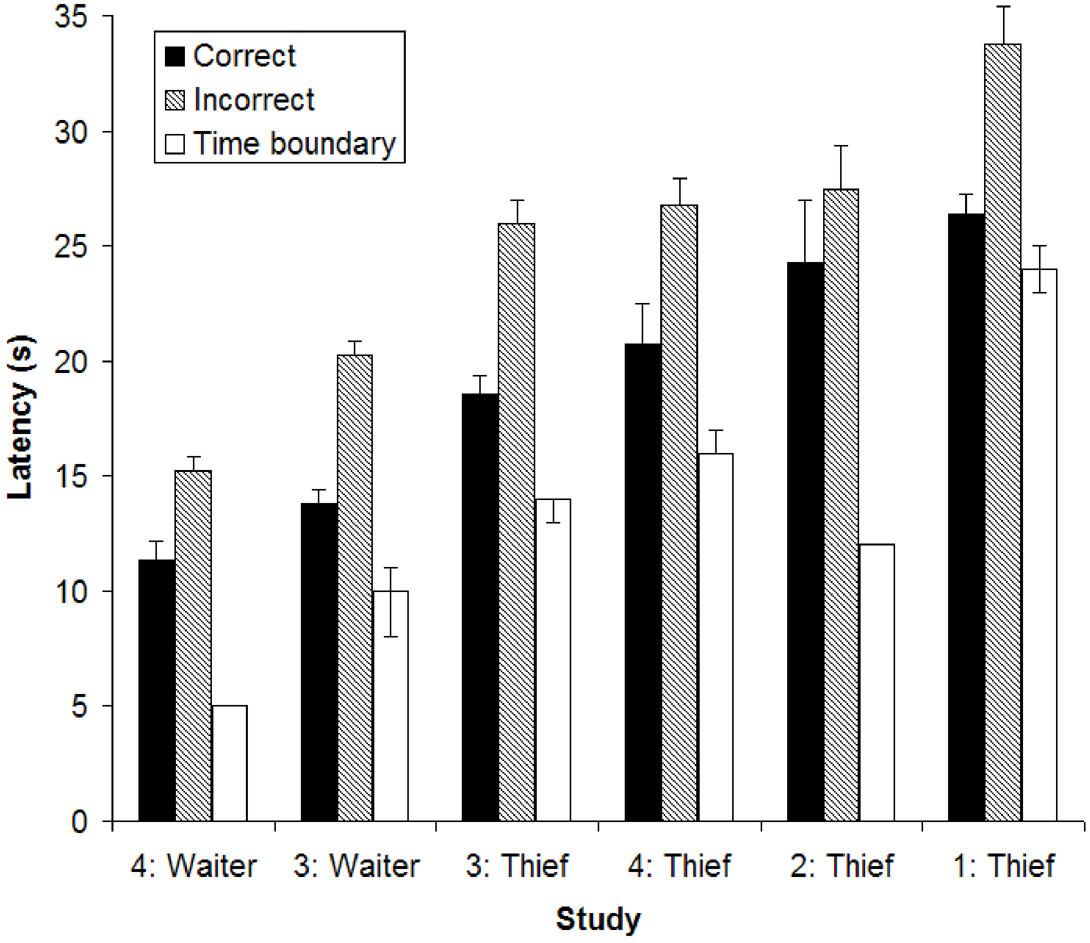

35
30
25
20
15
10
5
0
Latency (s)
■ Correct
▨ Incorrect
☐ Time boundary
4: Waiter
3: Waiter
3: Thief
4: Thief
2: Thief
1: Thief
Study

Figure 2