Predicting eyewitness identification accuracy with mock witness measures of lineup fairness: Quality of encoding interacts with lineup format

Colin Tredoux
ACSENT Laboratory, University of Cape Town, South Africa
colin.tredoux@uct.ac.za

Janat Fraser Parker
Department of Psychology, Florida International University, Miami, United States of America

David Nunez
ACSENT Laboratory, University of Cape Town, South Africa

After viewing a staged crime under varying levels of encoding optimality, eyewitnesses made identifications from lineups that were either presented to them as a simultaneous line of people, or individually, in sequence. Lineups did not contain the perpetrator of the staged crime, and were purposefully constructed to have varying levels of bias. There was an interaction between encoding optimality and type of lineup, with sequential-presentation lineups leading to more correct rejections than simultaneous-presentation lineups for moderate encoding optimality. Sequential-presentation lineups did not provide good protection against mistaken identifications in the two other conditions of encoding optimality. Measures of the fairness of the lineups significantly predicted identification rates for both simultaneous-presentation lineups and sequential-presentation lineups, regardless of encoding optimality.

Keywords: encoding optimality; eyewitness testimony; identification parades; lineup fairness; mock witnesses

Criminal convictions often depend on identifications made by eyewitnesses, but these identifications are unfortunately quite often mistaken. Legal systems all over the world have recognised this fact for a long time — as early as 1926, a commentator in the *South African Law Journal* noted '… mistaken identity is the most likely and common cause of miscarriages of justice …' ('Justice of the Peace', 1926, p. 287) — but more recently DNA testing has suggested that the scale of the problem is vast. The Innocence Project (www.

ISSN 0081-2464

innocenceproject.org) has revealed definitive DNA evidence of mistaken conviction of over 160 people in the USA, and most of these cases were decided on the basis of eyewitness evidence. In South Africa, in the case of the 'Eikenhof Three' (Tredoux & Chiroro, 2005), three African National Congress operatives were mistakenly convicted and sentenced to death (later commuted to life imprisonment) in 1993, on the basis of eyewitness evidence, and confessions made under duress. They were released from prison after serving five years.

Historically, lineup identifications have been critical to testing eyewitness evidence of identification (lineups are also known as 'identification parades' in South Africa and England). However, lineups can be biased against the suspect, or present witnesses with few plausible choices, and so methods to measure lineup fairness have been developed. Typically, lineups are evaluated by asking 'mock witnesses' who did not witness the crime to identify the suspect, based solely on the description of the suspect given by the eyewitness. Numerous lineup fairness measures based on the mock witness method have been recommended (Malpass & Devine, 1983; Wells, Leippe, & Ostrom, 1979), and arguments have been made as to the superiority of measures in terms of sensitivity, discriminability, and amenability to statistical inference (e.g., Brigham, Ready, & Spier, 1990; Malpass, Tredoux, & McQuiston-Surrett, 2007; Tredoux, 1998). The key indices are those of lineup bias, typically measured as a function of the proportion of mock witnesses who choose the suspect (who should not be able to choose the suspect if it is fair), and lineup size, typically measured as a function of the pattern of suspect and foil choices made by mock witnesses (which should be equally spread over lineup members if the lineup is fair). Thus, Wells et al. (1979) introduced a bias measure known as 'functional size', which is the inverse of the proportion of witnesses choosing the suspect, and Malpass (1981) and Tredoux (1998) both introduced measures of 'effective size', which attempt to index the departure of mock witness choices from a uniform pattern across lineup members.

There has recently been a flurry of renewed interest in the mock witness method from a research point of view (Corey, Malpass, & McQuiston, 1999; McQuiston & Malpass, 2002; Malpass et al., 2007). Two key questions have surfaced in this work: the question of 'process parallelism' first investigated by Corey et al. (1999), and the question of what might be termed 'outcome parallelism'. The question of process parallelism concerns the differences that underlie the acts of witnessing and 'mock-witnessing'. Corey et al. (1999) make a convincing argument that there must be substantial processing differences between mock and real witnesses, and the practice of using mock witnesses as analogues of real witnesses may be subject to some difficulties. Mock witnesses must rely solely upon the description of the suspect, while eyewitnesses rely on memory images. As well as this cognitive disparity, the mock witness task has typically used a forced-choice biased instruction whereas the eyewitness task has used an unbiased instruction; that is, mock witnesses are required to choose one of the lineup members, but real eyewitnesses are not. In addition, the mock witness task has typically used a simultaneous-presentation array (i.e., where the lineup members are presented at the same time) and Corey et al.

(1999) speculate that it may not be applicable to sequential-presentation lineups (i.e., where the lineup members are introduced to the witness one at a time, and are not seen as a collective), but a later study by McQuiston and Malpass (2002), suggests that the mock witness technique can, with circumspect application, be used to assess the fairness of simultaneous lineups. These differences in process suggest that eyewitness identifications may vary under certain conditions whereas mock-witness identifications will remain stable, resulting in a disparity in outcomes. Because the processes involved are different, mock-witness estimates may not be predictive of eyewitness choices.

The question of 'outcome parallelism' was recently the subject of a series of studies by Lindsay and colleagues (Lindsay, Smith, & Pryke, 1999; Smith, Lindsay, & Pryke, 2000, 2001). Here the issue is whether the information given by a mock witness procedure (as indexed by measures of lineup size and lineup bias), predict (or postdict) the outcomes of real witness lineups, particularly the rate of false positive identifications. If mock witness measures are of any use, they should predict the performance of real witnesses with reasonable accuracy. Thus, if a mock witness task produces a low functional size for a lineup (strongly suggesting a lineup biased against the suspect), but real witnesses never falsely implicate an innocent suspect in the same lineup, the measure of functional size fails to say anything useful about the lineup.

Lindsay et al. (1999) ran a number of empirical studies that gathered data concerning the relationship between mock witness measures and real witness error rates (particularly false positives). These studies led Lindsay et al. (1999) to make several conclusions depending upon whether the mode of lineup presentation was simultaneous or sequential. With simultaneous-presentation, perpetrator-absent lineups (lineups in which the perpetrators of the staged crimes were absent, intended to simulate the analogous situation where police have arrested an innocent suspect and put him or her in a lineup), they observed that lineup size failed to postdict identification decisions although various measures of lineup bias were effective predictors. On the other hand, with sequential-presentation, perpetrator-absent lineups, they were unable to find that lineup bias or lineup size measures postdicted false positive decisions.

However, these conclusions may be premature, and more empirical data would be of considerable help. Tredoux (1999) has argued this at some length, so we will merely repeat one of his criticisms, namely that Lindsay et al.'s (1999) conclusion is based on non-independent data. Lindsay et al. (1999) treated each of the lineup members in their perpetrator-absent lineups as suspects, and calculated lineup measures and suspect identification probabilities for each lineup member. This permitted them to calculate lineup fairness measures and identification rates for a total of 90 perpetrators, despite having used only 15 perpetrator-absent lineups. The consequence of this is that the estimates of effective size calculated by Lindsay et al. (1999) show a severe restriction of range, and there must be a concomitant reduction in the strength of the relationship between effective size and probability of mistaken suspect identification. The relationship between lineup size and probability of mistaken identification may turn out to be very different when the data collection procedure ensures statistically independent estimates of lineup size (and identification probability).

In a later study with simultaneous lineups, Smith et al. (2000) conducted a similar analysis, using proportion of mock witnesses' choices as a measure of lineup fairness. They discovered a significant (if somewhat weak) relationship between this index of lineup fairness and identification accuracy. They did not use a measure of lineup size as a predictor, though, and their method of calculating lineup fairness used non-independent data in much the same manner as the earlier study.

Since there are no data on the relationship between lineup fairness and identification behaviour other than those collected by Lindsay and colleagues (Lindsay et al., 1999; Smith et al., 2000), it is imperative that additional data should be collected. Further, we suspect that a number of moderating variables affect the relationship between lineup fairness and eyewitness performance, which are likely to attenuate correlations between lineup measures (both size and bias) and suspect identification probability. In particular, it is likely that the optimality of the encoding conditions experienced by real witnesses determines later performance of those witnesses on a lineup task. In a high optimality condition, the witness has a good memory of the perpetrator, and will be less likely to be mistaken about an innocent suspect. The lower the optimality of the encoding, the more likely that the witness will be mistaken, and the more likely that he or she will not be able to reject lineup foils in a perpetrator-absent lineup. Mock witnesses, on the other hand, do not experience a similar variation in encoding optimality: their performance will depend on the structure of the lineup, and the nature of the description given to them. A procedure using mock witnesses can only generate one estimate of functional or effective size, and it follows that this estimate cannot be perfectly correlated with false positive rates from real witnesses, since real witnesses perform differently across different viewing conditions.

A second thrust of our study was to explore conditions that might limit the superiority of the sequential-presentation lineup. Lindsay et al. (1999) and Levi (1999) argue that the use of sequential-presentation lineups does away to a large extent with the need for measures of lineup fairness, since sequential-presentation lineups protect against poor eyewitnesses. It is also argued that sequential-presentation lineups protect against poor eyewitnesses and poor lineups better than simultaneous-presentation lineups (Cutler & Penrod, 1988; Lindsay & Wells, 1985; Lindsay et al., 1991). It is not clear, though, that they protect *sufficiently* against poor witnesses and poor lineups. In particular, a recent meta-analysis (Steblay, Dysart, Fulero, & Lindsay, 2001) shows that sequential-presentation lineups produce 46 per cent mistaken rejections in perpetrator-present lineups, and at least 9 per cent mistaken identification of innocent suspects in perpetrator-absent lineups. Recent research questions the robustness of the 'sequential superiority' effects, and argues that the apparent superiority may simply be a change in response criterion (Meissner, Tredoux, Parker, & Maclin, 2005), or the failure to use appropriate counterbalancing (McQuiston-Surrett, Malpass, & Tredoux, 2006).

In order to test some of these ideas, we conducted a staged-crime eyewitness experiment. We manipulated eyewitness exposure to a videotaped crime and examined eyewitness identification accuracy in relation to the level of lineup bias. To increase the correspondence between mock witnesses and eyewitnesses, unbiased instructions with a ‘none-of-the-above’ alternative were used. We also examined lineup presentation, as research has emphasised the importance of sequential-presentation lineups in reducing false identifications (e.g., Cutler & Penrod, 1988; Lindsay & Wells, 1985), yet mock witness measures are typically derived from simultaneous-presentation lineups (false identifications occur when a witness chooses an innocent suspect in a lineup from which the perpetrator is absent, and correct rejections occur when a witness indicates that the perpetrator is absent from the same kind of lineup).

In fact, Lindsay et al. (1999) were unable to obtain reliable correlations between mock-witness identifications and eyewitness identifications from sequentially-presented lineups (but see McQuiston and Malpass, 2002, for a successful use of mock witnesses with sequential lineups).

We had a number of expectations and predictions. In the first place, we expected that our manipulation of encoding optimality would moderate false identification and correct rejection rates in perpetrator-absent lineups. Specifically, we expected the false identification rates to be higher and the correct rejection rates to be lower when encoding optimality was poorer, and vice versa.

In the second place, we expected that sequential-presentation lineups would again show their advantage over simultaneous-presentation lineups, producing fewer false identifications and more correct rejections. However, we expected this protection to be moderated by lineup bias — specifically, we expected that sequential-presentation lineups would protect better against false identification rates when the lineup was strongly biased, in line with findings reported by Lindsay et al. (1991).

METHOD

Materials

The materials include videotapes of a staged crime, and perpetrator-absent photo-lineups.

Videotapes of simulated crime

A staged theft of a microwave oven was videotaped for each of two white male perpetrators (21 years and 26 years). Three versions were created to obtain three levels of encoding optimality: (a) high was a slow-motion 130 s recording, (b) moderate was a normal speed 66 s recording, and (c) low was an edited normal speed 40 s recording. The ‘optimality of encoding’ manipulation was constructed by editing the videotape recordings so as to produce systematic differences in identification performance in a pilot sample of 21 participants.

All scenarios were shown in monochrome and were preceded by a 10 s view of a videotaped card showing the ostensible date of the event, and the building name, written in longhand, in an attempt to present the constructed scenarios as real events. The scenarios were recorded on a JVC colour video camera, and edited to expand or shrink the amount of time the perpetrator was seen. This was done by means of two VCR units, and verified by stopwatch afterwards.

Lineup construction

The high optimality tape was shown to 17 pilot research participants who were asked to describe the person whom they had just seen steal the microwave oven. They were asked to provide general details such as height, weight, age, and hair colour (light or dark), as well as more specific details, but to list only features or details about which they were fairly sure. A consensus description was generated for each perpetrator, using the characteristics reported by the majority of the participants. For perpetrator one, the description was 'A white male, approximately 1.8 m tall, medium to well built, approximately 80 kg, and in his late twenties. He has short, dark hair, and a broad face.' For perpetrator two, the description was 'A white male, approximately 1.7 m tall, slender to average build, between 65 and 70 kg, and in his early twenties. He has short medium/dark brown hair, and an oval-shaped face.' (Note that we used two perpetrators in order to improve generalisability of results, and do not separate our analyses according to perpetrator — indeed, there was no systematic difference between them, either as main effects, or as interaction effects).

Based upon these descriptions, foils were chosen for the lineups, using a match-to-description strategy (see Luus & Wells, 1991). Black and white photos of male faces from eight databases (from the USA, UK, and South Africa) were rated by three independent raters for similarity to the description on five (perpetrator one) or six (perpetrator two) characteristics. For low-bias lineups, foils rated the most similar to the description were chosen; for moderate-bias lineups, foils rated similar to the description on three characteristics were chosen; for high-bias lineups, foils rated similar to the description on one or two characteristics were chosen. Only perpetrator-absent lineups were constructed for use in the study, and, in order to generate these, a suspect-substitute was randomly chosen for each perpetrator from the set of photos in which all characteristics were rated the same as the perpetrator. The suspect-substitute was presented in all six lineup positions for each of the three levels of lineup bias, and for each perpetrator, resulting in 36 different lineups.

This bias manipulation was checked by calculating lineup bias measures of functional and effective size with a pilot sample of 26 mock witnesses. Specifically, we used Wells et al.'s (1979) measure of functional size, and Tredoux's (1998) measure of effective size (E). Functional size is calculated by taking the inverse of the proportion of mock witnesses who choose the perpetrator (and therefore varies between 1 and $\propto$; the lineup is said to be unbiased when functional size = k, the size of the lineup). E is calculated according to the formula reported in Tredoux (1998), and varies between 1 and k, the size of the lineup (the most desirable value of E is k, and this is achieved when choices are equally distributed across the lineup members). On the basis of this pilot study, changes in lineups were made keeping within the aforementioned parameters.

Mock witnesses

Three hundred and ninety-three first-year psychology students participated as mock witnesses (we do not take note of their race or gender in our analysis, as the mock witness task is not a memory task, and these attributes are not known to affect performance on the task). Witnesses were presented booklets informing them of the crime and giving them a description of the suspect (the aggregated description used in the construction of the lineups). They were informed that the perpetrator might be in the lineup or might not be there at all. They were to indicate who they thought was the accused person (see Wells & Bradfield, 1999 for a rationale for using 'accused') by circling the number beneath one of the pictures or circling the 'none-of- the-above' alternative. If they chose 'none-of-the-above' they were later informed that the suspect was actually present and that they should choose a photo. Thirty-nine participants chose the 'none of the above' alternative and 32 of these then made a forced choice, whereas seven failed to follow instructions to choose a photo. Since results did not differ across forced and free-choice conditions, we considered the results from the forced choice condition in our analysis.

Eyewitnesses

Participants and design

Two hundred and sixty second-year psychology students (mean age = 20 years; *SD* = 2.48; range = 18 to 42 years) participated as eyewitnesses to the crime. Seventy-nine per cent were female and 20% were male. The racial distribution was 59% white, 16% black, 16% coloured, 6% Indian, and 3% Asian. Optimality of encoding (low, moderate, or high) was factorially combined with lineup presentation (simultaneous or sequential), lineup bias (low, moderate, or high), and perpetrator (1 or 2). Observant readers will note that 41% of our sample was exposed to a perpetrator of a different race, and that this can be expected to affect the accuracy of subsequent identifications. This effect is known as the 'cross-race bias', is moderately small, and has been shown to be present in Southern Africa as well as many other countries (Chiroro & Valentine, 1995; Wright, Boyd, & Tredoux, 2003). However, analysis of the cross-race effect in this study showed that race groups did not differ in their rate of false identifications of the suspect (the key dependent variable), and so we do not consider race differences any further here.

Procedure

Twenty-four groups of students witnessed the crime simulation. Prior to the video, students were informed that the researchers were doing applied research in consultation with campus security, and that campus security wanted them to judge the quality of their fixed video-camera material. They were asked to view a piece of footage that was captured at the university early one morning. Each group saw a perpetrator steal a microwave oven in a high, moderate, or low optimality video. Students were then given written open-ended instructions asking them to describe the person who stole the microwave oven.

After describing the perpetrator, students went to one of two offices where computers displayed perpetrator-absent lineups to the students individually. At the computer, students were randomly assigned to one of 36 cells (sequential or simultaneous presentation; low, moderate, or high lineup bias; position one until six for suspect-substitute in lineup). Eyewitnesses completed the task interactively with a computer, but recorded their choices on response sheets. The simultaneous lineup was shown as a ‘contact sheet’ (or photo-spread) on the computer at a resolution of 800 × 600 pixels. Each face image was 120 × 150 pixels in size. The sequential lineup was shown one image at a time and the individual pictures were comparable in size to those shown in the simultaneous lineup. Witnesses started the program, and it responded by explaining the task, namely that they should choose the perpetrator, if present. They were warned that the perpetrator might not be there at all and that they should choose ‘none of the above’ if that were the case. For simultaneous-presentation lineups, the lineup was displayed and the witness wrote the appropriate number on the response sheet accompanied by a seven-point confidence rating, or wrote ‘none of the above’, accompanied by a seven-point confidence rating. The confidence scale ranged from ‘very uncertain’ to ‘very certain’. For sequential-presentation lineups, witnesses were told they would see one photo at a time and that they should choose the perpetrator, if present, or ‘none of the above’, if not present. As each photo was displayed the witnesses wrote the appropriate number on the response sheet and checked ‘yes’ or ‘no’. There were eight numbered lines although only six photos were presented. After the sixth photo, the students were instructed on the computer to go to the bottom of the sheet and decide whether ‘none of the above’ was the decision they wished to make.

RESULTS

Mock witness results

Calculations of Tredoux’s (1998) effective size (*E*) and Wells et al.’s (1979) functional size were derived from the mock witness identifications to determine if our method of constructing lineups resulted in different levels of structural bias. The average values (across both perpetrators) of *E* and of the functional size were 3.80 and 3.58, respectively, for low bias lineups, 2.53 and 1.67, respectively, for moderate bias lineups, and 1.57 and 1.26, respectively, for high bias lineups.

Eyewitness results

Overall, 51.2 per cent of the research participants made the correct choice, namely a correct rejection (since all lineups were perpetrator-absent). Loglinear analyses were carried out on correct rejections and false identifications (identification of suspect-substitute) as a function of lineup presentation, lineup bias, and optimality of encoding. Table 1 shows the mean number of correct rejections and false identifications as a function of lineup presentation, lineup bias, and optimality of encoding.

Table 1. Percentage of correct rejections and false identifications as a function of lineup presentation, lineup bias, and optimality of encoding

		Lineup choice					
		Correct rejections			False identifications		
Lineup presentation	Lineup bias	Low optimality	Moderate optimality	High optimality	Low optimality	Moderate optimality	High optimality
	Low	40.0 (15)	37.5 (16)	46.2 (13)	33.3 (15)	25.0 (16)	15.4 (13)
Simultaneous	Moderate	61.5 (13)	15.4 (13)	60.0 (15)	30.8 (13)	61.5 (13)	33.3 (15)
	High	47.1 (17)	25.0 (12)	75.0 (12)	47.1 (17)	58.3 (12)	16.7 (12)
	Low	76.9 (13)	60.0 (15)	60.0 (15)	15.4 (13)	26.7 (15)	20.0 (15)
Sequential	Moderate	55.6 (18)	78.6 (14)	53.8 (13)	33.3 (18)	21.4 (14)	15.4 (13)
	High	47.1 (17)	46.7 (15)	35.7 (14)	29.4 (17)	40.0 (15)	57.1 (14)

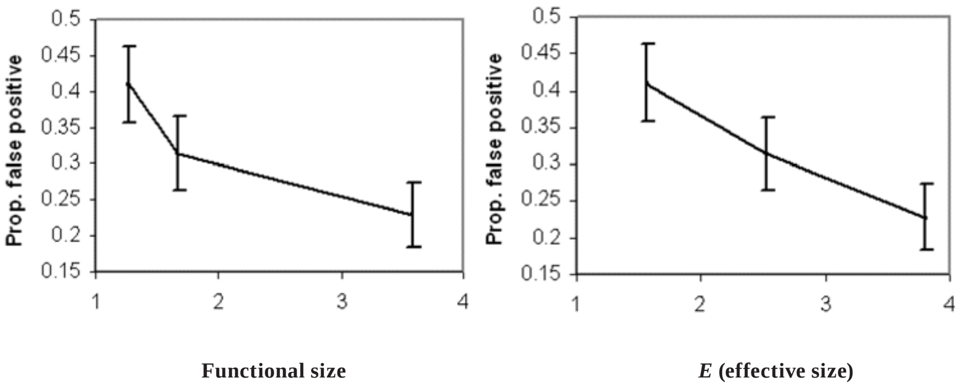
Note: Cell counts are shown in parentheses. Total $N = 260$

For correct rejections, lineup presentation was not significant, χ^2 (1, N = 260) = 3.58, p = 0.059, although the effect was in the expected direction. Although there was no main effect of optimality of encoding, χ^2 (2, N = 260) = 1.99, p = 0.37, there was an interaction of optimality of encoding with lineup presentation, χ^2 (2, N = 260) = 8.79, p = 0.01, showing that sequential-presentation lineups (61.4%) surpassed simultaneous-presentation lineups (26.8%) in correct rejections at the moderate level of encoding optimality, χ^2 (1, N = 85) = 10.24, $p < 0.005$, but sequential-presentation lineups did not differ from simultaneous-presentation lineups at the low level of optimality (58.3% vs. 48.9%, respectively), χ^2 (1, N = 93) = 0.83, $p > 0.05$, nor at the high level of optimality (50% vs. 60% respectively), χ^2 (1, N = 82) = 0.83, $p > 0.05$.

A similar loglinear analysis was also carried out on false identifications as a function of lineup presentation, lineup bias, and optimality of encoding. The only significant finding for false identifications was a lineup bias effect, χ^2 (2, N = 260) = 7.22, $p < 0.05$, showing more false identifications in the high bias condition (41.4%) than in the low bias condition (23%), χ^2 (1, N = 174) = 6.74, $p < 0.01$. There were no differences between the low bias condition (23%) and the moderate bias condition (32.6%) nor the moderate bias condition (32.65) and the high bias condition (41.4%), $p > 0.05$. Optimality of encoding was not a significant factor for false identifications; the high, moderate, and low optimality conditions (27%, 37%, and 32%, respectively) did not differ, χ^2 (2, N = 260) = 2.46, $p >$ 0.29. Likewise, there was no effect of lineup presentation, χ^2 (1, N = 260) = 1.55, $p > 0.21$.

Finally, a loglinear analysis was carried out on foil identifications (i.e., identifications of members of the lineup other than the suspect) as a function of type of lineup, lineup bias, and optimality of encoding, but there were no significant findings.

An important result, given the rationale for the present study, was the relationship between lineup measures (bias and size), as derived from the mock witness task, and the rate of false identifications by eyewitnesses. There was a clear, decreasing, relationship between measures of E, functional size, and the percentage of false suspect identifications made by eyewitnesses. Figure 1 shows a plot of the relationships, averaged over simultaneous and sequential lineups. A breakdown of these relationships by lineup type (not shown in the figure) revealed the same basic pattern, albeit not quite as strong as that shown in the figure. Goodness of fit tests were conducted in order to assess the predictiveness of the monotone relationships shown in Figure 1. Ratios of the three functional sizes were used to generate fitted values, and the resulting χ^2 value gave no reason to reject the fit, χ^2 (2, N = 260) = 3.26, $p > 0.19$. A similar test on the values of E also gave no reason to reject the fit, χ^2 (2, N = 260) = 1.26, $p > 0.53$.



Note: standard error bars extend above and below points.

Figure 1. The relationship between two (mock witness) lineup measures and false eyewitness identifications

As 41 per cent of the students were tested in one office with the instructor of their class as experimenter and 59 per cent were tested in a second similar office with a graduate assistant as experimenter, experimenter was entered as a variable in analyses of correct rejections and false identifications. Surprisingly, a main effect of experimenter emerged in both the analysis of correct rejections, χ^2 (1, N = 260) = 7.40, $p < 0.01$ and false identifications, χ^2

(1, $N = 260$) = 4.66, $p < 0.05$, but all previous findings were still significant when this was factored into appropriate models. For correct rejections, the effect of lineup presentation was not significant, χ^2 (1, $N = 260$) = 3.21, $p = 0.07$ and the interaction of encoding optimality with lineup presentation was also still evident, χ^2 (1, $N = 260$) = 9.81, $p = 0.007$. There was an interaction of lineup presentation with experimenter, χ^2 (1, $N = 260$) = 5.55, $p = 0.02$, showing greater correct rejections for sequential (72.4%) over simultaneous presentation (49%) for the faculty experimenter, χ^2 (1, $N = 107$) = 6.17, $p = 0.01$ but not for the graduate student experimenter (44.7% vs. 42.9%), χ^2 (1, $N = 153$) = 0.06, $p = 0.82$. This interaction also showed that correct rejections in simultaneous-presentation lineups did not differ across faculty (49%) and graduate student (42.9%) experimenters, χ^2 (1, $N = 126$) = .45, $p = 0.50$, but more correct rejections were evident in sequential-presentation lineups in the presence of the faculty experimenter (72.4%) than in the presence of the graduate student (44.7%), χ^2 (1, $N = 134$) = 10.26, $p = 0.001$.

Although an experimenter effect also emerged in the analysis of false identifications, the main effect of lineup bias was still significant, χ^2 (1, $N = 260$) = 7.23, $p = 0.03$. Thus, although the experimenter effect was evident for both major dependent variables, the original findings were still very much in evidence.

DISCUSSION

The results of the present experiment emphasise that we must be vigilant in determining and reporting encoding conditions in laboratory experiments and real-world events. In the present study, there was evidence that the typical superiority of sequential-presentation over simultaneous-presentation lineups in terms of correct rejections was only evident in the moderate optimality of encoding condition. There was no advantage in either the low optimality condition, or the high optimality condition. The moderate optimality condition is most comparable to the characteristic level of exposure in an experimental study, namely a normal speed videotape of an event with moderate duration. Although the direction of the difference was similar in the low optimality of encoding condition (normal speed videotape of an event but shorter duration), the difference across lineup types did not approach significance. The high optimality condition, on the other hand, attempted to increase the opportunity for accurate observation (by slowing the speed of the video, thus increasing the duration of the event), but in fact did not result in an advantage for sequential lineups. This latter condition certainly increased the opportunity to encode the event, but, in retrospect, is not a simple or ecologically ‘real’ manipulation of duration. Regardless, it is clear that there are presentation conditions that do not manifest the typical superiority of sequential-presentation over simultaneous-presentation lineups. As Steblay et al.’s (2001) meta-analysis suggests, it appears then that sequential-presentation lineups do not offer ‘blanket’ protection. Other recent studies come to a similar conclusion (McQuiston-Surrett et al., 2006; Meissner et al., 2005).

In the light of these differential findings as a function of optimality of encoding, additional studies should be done to operationalise the optimality of encoding variable. There are various ways to operationalise optimality, such as duration of exposure to the perpetrator (e.g., total amount of time) or degradation of the quality of the image (e.g., a low-resolution pixellation) to produce poorer encoding conditions. Some manipulations may be more ecologically valid than others, but theoretically all should be explored.

A rather unexpected result in the present study was the presence of an experimenter effect, particularly with sequential-presentation lineups. We specifically tried to control for this possibility by implementing the lineup test as an ostensibly experimenter-free computerised task. The only contact experimenters had with participants was to meet them at the door of the laboratory, guide them to a seat in front of the computer, and explain the task to them. Experimenters did not interact with participants during the task. One of the experimenters, however, was a permanent faculty member who had lectured introductory psychology students on eyewitness research, whereas the other was a graduate assistant. Since the former had stressed how easily eyewitnesses can make mistaken identifications, it is possible that the second-year psychology students who completed the lineup task under his observation remembered this from the previous year of study, and were much more cautious (they were indeed 18 per cent more likely to make a correct rejection).

This result is consistent with Phillips, McAuliff, Kovera, and Cutler's (1999) findings that sequential-presentation lineups are more sensitive to experimenter bias. Phillips et al. observed that witnesses were more likely to make a false identification in sequential-presentation lineups when an observer was present and the lineup administrator was not blind to the identity of the suspect (single-blind procedure). In the present study, the experimenters can actually be considered comparable to the observers in the Phillips et al. study (in fact the observers in Phillips et al. were the actual experimenters and participants were used as lineup administrators). It appears that demand characteristics (Rosenthal & Rosnow, 1991) may have played a role in both studies. In the Phillips et al. study, the presence of the experimenter-observer may have increased the pressure felt by both the lineup administrator and the eyewitness to provide a positive identification. On the other hand, in the present study, students may have wanted to yield to the authority of their instructor and also win his approval by demonstrating caution. Interestingly, these effects only seem to manifest themselves in sequential-presentation lineups.

The method of constructing lineups by choosing distractors that differed in the number of characteristics that matched the description of the perpetrator was effective in creating different levels of structural bias as determined by established measures of lineup bias and lineup size. The false identification results demonstrate once more the concurrent validity between measures of lineup bias and manipulations of similarity. This is consistent with the findings of Lindsay and Wells (1980), in which more extreme variations in lineup similarity were manipulated. These findings were evident for both simultaneous- and sequential-presentation lineups.

Lineup measures in this study, as derived from a mock witness task, were predictive of later eyewitness performance for both simultaneous- and sequential-presentation lineups. This was true for an index of lineup bias (functional size) and for an index of lineup size (*E*). This is consistent with Smith et al.'s (2000) finding that a lineup fairness measure was one of the best postdictors of accuracy, at least in simultaneous-presentation lineups. However, it is at variance with their argument that lineup measures derived from mock witness tasks do not have adequate criterion validity, particularly for sequential-presentation lineups. The data from the present study suggest that lineups that have low functional size and low effective size, be they simultaneous or sequential, lead to more false identifications of innocent suspects than those that have higher functional size and higher effective size.

The resurgence of interest in the mock witness method (an entire issue of *Applied Cognitive Psychology* was devoted to the topic in 1999) appears to be justified. Measures of lineup fairness derived from the method proved to be predictive of eyewitness identification accuracy. This was true, on the one hand, for measures of lineup bias and lineup size. It does not appear to be true, on the other hand, that sequential-presentation lineups protect against false identifications so well that they effectively eliminate the need to measure lineup fairness with mock witnesses. In the present study, we found a significant and troubling number of false identifications in sequential-presentation lineups, consistent with Steblay et al.'s (2001) meta-analysis.

To summarise: We hypothesised that encoding optimality would moderate false identification and correct rejection rates in perpetrator-absent lineups and that sequential-presentation lineups would lead to fewer false identifications and more correct rejections than simultaneous-presentation lineups, particularly under conditions of strong bias or unfairness. We also hypothesised that mock witness measures of lineup fairness would predict rates of false identifications.

We found that optimality of encoding interacted with lineup modality, such that sequential-presentation lineups encouraged more correct rejections of perpetrator-absent lineups, but only when encoding optimality was moderate. We did not find a general advantage for sequential-presentation lineups over simultaneous-presentation lineups, nor did we find a specific advantage when lineups were strongly unfair, counter to our predictions.

Mock witness measures of lineup fairness, on the other hand, proved to be predictive of rate of false identifications by eyewitnesses in both simultaneous- and sequential-presentation perpetrator-absent lineups.

CONCLUSIONS

The results of the current study are of particular importance in two respects. First, they address the issue of which type of lineup presentation is superior. It appears that sequential-presentation lineups are superior to simultaneous-presentation lineups only in carefully circumscribed situations, such as eyewitness events of moderate encoding possibilities. Further, sequential-presentation lineups appear to be more susceptible to experimenter bias, even in situations that presumably should be minimising such bias (such as computer presentation of photos). Some researchers have strongly advocated the implementation of the sequential lineup in police practice, and have achieved this goal in some jurisdictions in the United States (see Kloberdanz, 2005; Levi & Lindsay, 2001; Lindsay, 1999). We believe that this advocacy has been premature, and the results of this study, as well as those from other recent studies (McQuiston-Surrett et al., 2006; Meissner et al., 2005) bear this contention out.

Second, the issue of parallelism between eyewitness identifications and mock witness identifications was explored. Although the processes are quite different in terms of both cognitive and procedural properties, the results of this study provide confirmation for using the mock witness procedure to evaluate fairness of lineups. There was a clear monotonic relationship between false identifications and lineup fairness measures as determined by two different measures of bias.

REFERENCES

Brigham, J. C., Ready, D. J., & Spier, S. A. (1990). Standards for evaluating the fairness of photograph lineups. *Basic and Applied Psychology, 11*(2), 149–163.

Chiroro, P., & Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. *Quarterly Journal of Experimental Psychology, 48A*, 879–894.

Corey, D., Malpass, R. S., & McQuiston, D. E. (1999). Parallelism in eyewitness and mock witness identifications. *Applied Cognitive Psychology, 13,* S41–S58.

Cutler, B. L., & Penrod, S. D. (1988). Improving the reliability of eyewitness identification: Lineup construction and presentation. *Journal of Applied Psychology, 73*(2), 281–290.

'Justice of the Peace' (1926). Identification. *South African Law Journal, 43,* 288.

Kloberdanz, R. (2005, March 21). Building a better lineup. *Time*. Retrieved November 21, 2006, from www.time.com/time/magazine/article/0,9171,1037673,00.html

Levi, A. M. (1999). An honourable discharge for line-up fairness measurement. *Applied Cognitive Psychology, 13,* S121–S124.

Levi, A. M., & Lindsay, R. C. L. (2001). Lineup and photospread procedures: Issues concerning policy recommendations. *Psychology, Public Policy, and Law, 7,* 776–790.

Lindsay, R. C. L. (1999). Applying applied research: Selling the sequential line-up. *Applied Cognitive Psychology, 13*, 219–225.

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*(3), 556–564.

Lindsay, R. C. L., Lea, J. A., Nosworthy, G. L., Fulford, J. A., Hector, J., leVan, V., &

Seabrook, C. (1991). Biased lineups: Sequential presentation reduces the problem. *Journal of Applied Psychology, 76*, 796–802.

Lindsay, R. C. L., Smith, S. M., & Pryke, S. (1999). Measures of lineup fairness: Do they postdict identification accuracy? *Applied Cognitive Psychology, 13,* S93–S107.

Luus, C. A. E., & Wells, G. L. (1991). Eyewitness identification and the selection of distractors for lineups. *Law and Human Behavior, 15,* 43–57.

Malpass, R. S. (1981). Effective size and defendant bias in eyewitness identification lineups. *Law and Human Behavior, 5*(4), 299–309.

Malpass, R. S., & Devine, P. G. (1983). Measuring the fairness of eyewitness identification in lineups. In R. S. Malpass & P. G. Devine (Eds.), *Evaluating witness evidence* (pp. 81–102). London: Wiley.

Malpass, R. S., Tredoux, C. G., & McQuiston-Surrett, D. E. (2007). Lineup construction and lineup fairness. In R. Lindsay, D. F., Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of eyewitness psychology, vol. 2: Memory for people* (pp. 155–178). Mahwah, NJ: Erlbaum.

McQuiston, D. E., & Malpass, R. S. (2002). Validity of the mockwitness paradigm: Testing the assumptions. *Law and Human Behavior, 26*(4), 439–453.

McQuiston-Surrett, D. E., Malpass, R. S., & Tredoux, C. G. (2006). Sequential vs. simultaneous lineups: A review of methods, data, and theory. *Psychology, Public Policy and Law, 12*(2), 137–169.

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

Phillips, M. R., McAuliff, B. D., Kovera, M. B., & Cutler, B. L. (1999). Double-blind photoarray administration as a safeguard against investigator bias. *Journal of Applied Psychology, 84,* 940–951.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis*. New York: McGraw Hill.

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. (2001). Postdictors of eyewitness errors: Can false identifications be diagnosed in the cross race situation? *Psychology, Public Policy, and Law, 5*, 153–169.

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 25,* 459–474.

Tredoux, C. (1998). Statistical inference on measures of lineup fairness. *Law and Human Behavior, 22,* 217–237.

Tredoux, C. (1999). Statistical considerations when determining measures of lineup size and lineup bias. *Applied Cognitive Psychology, 13,* S9–S26.

Tredoux, C. G., & Chiroro, P. (2005). Eyewitness testimony. In C. G. Tredoux, D. H.

Foster, A. Allan, A. Cohen, & D. Wassenaar (Eds.), *Psychology and law* (pp. 194–226). Cape Town: Juta.

Wells, G. L., & Bradfield, A. L. (1999). Measuring the goodness of lineups. Parameter estimation, question effects, and limits to the mock witness paradigm. *Applied Cognitive Psychology, 13,* S27–S39.

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.

Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2003). Inter-racial contact and the own race bias for face recognition in South Africa and England. *Applied Cognitive Psychology, 17*, 365–373.