Running head: Criminal Face Bias in Police Lineups

An Examination of Criminal Face Bias in a Random Sample of Police Lineups

Heather D. Flowe and Joyce E. Humphries

University of Leicester

Address ms. correspondence to:

Heather D. Flowe
hf49@le.ac.uk
School of Psychology
Forensic Section
University of Leicester
106 New Walk
Leicester, UK
LE1 7EA

Abstract

Faces with a stereotypic criminal appearance are remembered better and identified more often than other faces according to past research. In the present project, a random sample of police lineups was evaluated using the mock witness paradigm to determine whether criminal appearance was associated with lineup choices. In Study 1, mock witnesses were either provided with a description of the culprit or they were not. Participants also self-reported why they had selected a given face. In Study 2, the line-up faces were rated with respect to criminal appearance, distinctiveness, typicality, and physical similarity. Criminal appearance was the primary reason self reported for face selection in the no description condition. Mock witness choices in the no description condition were associated with only criminal appearance. When provided with a description, mock witnesses based their choice on the description. These findings are discussed in relation to lineup fairness.

An Examination of Criminal Face Bias in a Random Sample of Police Lineups

Faces that have a stereotypic criminal appearance are remembered better and identified more often than other faces, a phenomenon known as criminal face bias. Criminal appearance has been found to influence memory encoding, such that faces rated relatively high on criminality are remembered better and hence identified more often than their counterparts (MacLin & MacLin, 2004; c.f. Yarmey, 1993). Moreover, people may have stereotypes concerning the physical appearance of specific types of criminals, such as rapists and murderers (e.g. Bull & Rumsey, 1988; Goldstein, Chance, & Gilbert, 1984; MacLin & Herrera, 2006; MacLin, Waack, & Peterson, 2003; Shoemaker, South, & Lowe, 1973). Research has further found that people can be predisposed toward selecting certain types of faces from a lineup when they rely on these stereotypes (MacLin, Malpass, & Herrera, 2001). Thus, extant research suggests that criminal appearance can influence face discriminability and judgment bias.

The present study examined the potential extra-memory influence of criminal appearance on eyewitness identifications. We extend previous research by examining whether criminal appearance would be related to mock witness identifications from lineups that were randomly sampled from actual police cases. Doob and Kirshenbaum (1973) devised the mock witness paradigm, which is a widely used method for assessing lineup fairness. Mock witnesses—or research participants who did not witness the crime and who are therefore blind to the actual appearance of the culprit, are presented with a lineup and a physical description of the culprit. They are asked to identify the person who most closely matches the description. If the rate of choosing the suspect is above chance expectation when mock witnesses are armed with only a description, this suggests that the lineup has not been constructed in a fair manner. In the first

study we report, mock witnesses were presented with the police lineups and asked to determine which member was the police suspect. Mock witnesses also provided the reason for the identification they made in each lineup. Responses were analysed to determine how often criminal appearance was reportedly used to make an identification. In Study 2, participants rated each of the lineup faces with respect to criminality, distinctiveness, typicality, and lineup member similarity. The ratings were correlated with the mock witness identification rates obtained in Study 1. Using these methods, the following research questions were addressed: Would mock witnesses select a face based on their preconceived notions of what a criminal looks like? Does criminal appearance contribute to mock witness identifications once lineup member similarity and face distinctiveness are taken into account?

Criminal face bias could exert an extra-memory influence on eyewitness identifications on at least two levels. First, a judgment bias on the part of the eyewitness may lead to the identification of the face that most closely matches the physical appearance of a stereotypic criminal. If perchance the most criminal-looking person is the suspect, then eyewitnesses may be biased toward picking the suspect. Eyewitnesses could also be biased away from picking the suspect if there is someone more criminal-looking in the lineup (MacLin et al., 2001). Second, criminal appearance might affect eyewitness identifications because the person constructing the lineup selects foils, either consciously, or inadvertently, that are less criminal in physical appearance compared to the suspect. Thus, the suspect may "pop out" in a criminal appearance biased lineup because eyewitnesses have a pre-existing bias to choose the most criminal-looking face or because the suspect is simply different from the foils in criminal appearance.

Criminal face stereotypes could be based on features that arise from the physiogamy of the face (see Lombroso, 2005) or from more ephemeral facial characteristics that arose at the time

when the mug shot photograph was taken. Features that people associate with criminality include having long or shaggy dark hair, tattoos, beady eyes, pock marks and scars (MacLin & Herrera, 2006). Physical unattractiveness has also been associated with criminality such that physically unattractive people are rated as being more likely to commit criminal acts than attractive people (Bull, 1979; MacLin & MacLin, 2004; Mocan & Tekin, 2005; Saladin, Saper, & Breen, 1988). The specific elements associated with criminal face stereotypes, however remain unclear (MacLin & Herrera, 2006). More ephemeral characteristics of the suspect could include negative emotional expressions, lack of eye contact, or a disheveled appearance when the mug shot photograph was taken. Arrestees could also be under the influence of drugs and/or alcohol, which could affect, say, whether their eyes are bloodshot and glazed over as well as their emotional expression. In the current project, we utilised lineups in which the police had selected foils (i.e., distractors) from a large mug shot database that contained several thousand faces. All of the persons in the lineups had at one time or another been suspected of having committed a crime. Consequently, if criminal-biased lineups were found, the bias arose because of the methods that were used to select the foils in constructing the lineup, not for other reasons.

Criminal face bias has been previously examined in the mock witness paradigm. MacLin et al. (2001) constructed lineups using mug shots from a police database. The lineup members were chosen based on their match to a physical description. Independent raters evaluated the lineups to determine which members best fit the description and which appeared the most criminal-looking. The lineups were then presented to mock witnesses, who received criminal information or physical appearance information about the suspect, or both types of information. They found that when mock witnesses were given physical information or both physical and criminal

information, the member who best fit the description was selected most often. When mock witnesses were given criminal information, however, choices were directed toward the most criminal-looking member.

We followed the lead of MacLin and colleagues and used lineups composed of actual police mug shots to increase the external validity of the face stimuli. In the present study, however, the stimuli were lineups that the police had arranged for use in actual criminal cases. In so doing, we were able to examine the extent to which criminal appearance varied across the members of a lineup that was constructed by the police. Real world lineups may substantially differ from those used in the laboratory. To illustrate, Flowe et al. (2010) found that about a third of real world criminal suspects have distinctive features, such as scars, gold-capped teeth and facial tattoos, whereas face stimuli employed in eyewitness identification laboratory studies do not have these types of distinctive features. As such, it may be more ecologically valid to examine the relationship between criminal appearance and identifications using real world lineups.

The relationship between criminal appearance and mock witness identifications was expected to vary depending on whether a description of the culprit was provided. Brigham, Meissner, and Wasserman (1999) found levels of suspect bias, or the tendency to choose the suspect at above chance levels, varied depending on whether or not mock witnesses had a description of the culprit. Suspect bias was also found to vary for a given lineup depending on the specific features that were mentioned in the culprit's description. We expected that criminal appearance would be associated with identification decisions when mock witnesses had no other information on which to base their identifications. To examine this possibility, Study 1 varied whether mock witnesses were given a description of the culprit. We also elected to inform participants that they were viewing actual police lineups. Wells and Bradfield (1999) found suspect bias rates varied for

mock witnesses given a description of the culprit depending on the specific question posed.

viewing actual police lineups. Mock witnesses in Study 1 were also asked to provide an

explanation for each of their identifications. The reasons were analysed to determine how often

criminality was self-reported as the primary reason behind an identification.

In Study 2, criminal appearance was objectively measured by having participants rate each of

the lineup faces. One purpose in so doing was to determine how often the suspect would be

ranked the highest among the lineup members with respect to criminal appearance. Other groups

of participants rated the distinctiveness and typicality of the faces, as well as the physical

similarity of the lineup members. Researchers often measure the similarity of the lineup

members to compose lineups for use in laboratory studies (Flowe et al., 2010). Differences

across lineup members in criminal appearance, however, may not be discernable with physical

similarity or distinctiveness ratings. This may be because similarity and distinctiveness are

related to the physical features of faces, whereas criminal appearance is related to other qualities

of faces, such as emotional expression. We hypothesised that criminal appearance is a dimension

of faces that is separable from physical similarity and distinctiveness. This hypothesis was tested

by examining whether the mock witness results from Study 1 were correlated with criminal

appearance ratings once the distinctiveness and physical similarity ratings of the lineup members

were taken into account.

STUDY 1

Method

Participants

A total of 48 undergraduates (56% female) participated as mock witnesses; 23 people were

randomly assigned to participate in the description condition and 25 were assigned to the no

description condition. The mean age of the mock witnesses was 19.61 (SD = 1.70) years. The majority of participants indicated they were either Caucasian (46%) or Asian[1] (37%).

Materials

Criminal cases (robbery, theft, and assault) were randomly sampled from police arrest files and examined to determine whether they met the study's inclusion criteria, which were as follows: 1) a six person photographic lineup had been employed in the case, 2) a clear photocopy of the lineup was contained within the case file, 3) all of the lineup photographs were mug shots that had been taken from the police arrest database, and 4) at least one eyewitness to the crime provided a description of the perpetrator. Cases were sampled until a total of 11 lineups meeting these criteria had been located.

The charges were dropped against 5 of the suspects, whereas for the remaining 6, the suspect pleaded guilty. The race of the suspect, as described in the police record, was White in 4 of the lineups, Black in 4 of the lineups, and Hispanic in 3 of the lineups. The lineups were composed of black and white photographs and each contained six persons presented in a 3 X 2 array. The actual lineup as it appeared to the eyewitness(es) in the case was utilised (e.g., the position of the lineup members, the focal distance of each of the faces away from the camera, and the physical size of the photographs, etc. was maintained). The photographs in each lineup were numbered 1 through 6.

Eyewitness descriptions of the perpetrator were obtained from the Crime Incident Report, which is a routine form the police complete at the crime scene. The police recorded eyewitness descriptions on the Crime Incident Report using standard feature checklist. The checklist includes physical characteristics (e.g., height, weight, age, race, hairstyle, and eye colour) as well other traits (e.g., the perpetrator's temperament, cleanliness, sound of voice). If more than one

eyewitness provided a description in the case, we randomly selected one of them for use in the present study. A single description was utilised because the number of eyewitnesses providing descriptions varied across the 11 cases. We could have elected to use the modal eyewitness description in the cases that had multiple eyewitnesses. However, so doing may have resulted in these cases either having a greater number of descriptors and/or more correct descriptors compared to the cases that had only a single eyewitness. As a consequence, the ability of mock witnesses to identify the suspect might have varied across cases that had single versus multiple eyewitnesses. To control for this possibility, a single eyewitness description was randomly selected from cases that had multiple eyewitnesses for the present study. The eyewitness reported an average of 6 features (about 5 of which were physical descriptors and 1 of which was another type of trait) across the lineups.

Design and Procedure

Participants received 11 lineups, a response sheet, and written instructions. Lineup presentation order was randomly determined for every participant. Participants were randomly assigned to one of two mock witness conditions: In the description condition, each lineup was accompanied by a written physical description of the culprit, whereas in the no description condition, physical descriptions were not provided.

Participants in both description conditions were verbally told before the exercise commenced that they were going to evaluate criminal lineups that had been used in actual police cases. The lineups were to be evaluated in the order given. Participants were asked to determine for each of the lineups which of the persons they thought was the police suspect. In the description condition, the mock witnesses were further told that each of the lineups was accompanied by an eyewitness' description of the culprit; they were asked to use the description in making their

decision.

After the mock witnesses indicated their choice on the response sheet, they reported in a free response format the reason(s) why they had selected that particular member. No examples of particular reasons that could be given were provided, as we were interested in the number of people who would spontaneously report having used criminality as one of their criteria for having chosen the lineup member. Additionally, in the description condition, participants were asked whether they had used the eyewitness description in making their choice. An opportunity to ask questions was provided, and then participants were left alone in a room to complete their packet.

Measures

Identification Rates. For every lineup, the distribution of mock witness choices across the lineup members was determined. Mock witness choice distributions were aggregated separately for the description and no description conditions.

Mock Witness Self-Reports of Decision Processes. The mock witness free response data were coded by two people. Six categories (criminal appearance of the person, guilty appearance of the person, the person appearing physically different from the others, the "look" the person had in his eyes, whether the person was smiling or not, and other types of emotional expressions the person had on his face) could be distinguished, accounting for 70% of the mock witness selfreports.

Agreement between coders for the initial coding of the free responses for criminality (mean Kappa coefficient = 0.94, range: 0.81-1.00 across lineups) and guilt (mean Kappa coefficient = 0.93, range: 0.88-1.00 across lineups) was high. The data set was finalised after coders resolved disagreements through discussion.

Lineup Fairness Measures. Two measures of lineup bias were computed: Suspect bias was measured using Doob and Kirshenbaum's (1973) method, whereby the rate of picking the suspect was compared against chance expectation (.17) as a measure of the extent to which mock witnesses were biased toward picking the suspect. Additionally, following Wells, Leippe, and Ostrom (1979), functional size was computed for every lineup by dividing the total number of mock witnesses by the number of mock witnesses that selected the suspect. One measure of lineup size was computed for every lineup: Tredoux' E (Tredoux, 1998), which is a measure of the effective size of a lineup that takes into account the distribution of mock witness choices across lineup members. Suspect bias, functional size, and Tredoux' E were computed separately for mock witness data in the description and no description conditions.

Results

Preliminary Analyses

Identification rates for each face across the description and no description conditions were statistically independent (r = .01, p = .92). These results indicate that the rate of choosing a given face varied depending on whether mock witnesses had a description of the culprit. Additionally, when mock witnesses in the description condition were asked whether they made the identification based on the eyewitness' description of the culprit, the answer was "yes" for 69% in any of the lineups ($\bar{r}$ =.02, SDr = .06, r range: -.09 to .08). These results indicate that the order in which participants evaluated a given lineup was not related to mock witness choices; therefore, the order in which participants evaluated the lineups will not be discussed any further. Additionally, the race of the lineup members did not influence any of the measures. If anything, the raters (whether they were White or Asian) seemed to have a slight tendency toward viewing the White compared to the Black and Hispanic lineup members as less similar to one another and

more criminal in appearance; these differences, however, were not statistically significant. Therefore, race will not be discussed any further.

Was the suspect chosen more often if a description was given?

Table 1 provides the lineup bias and lineup size measures in the description and no description conditions. Suspect identification rates did not significantly change when mock witnesses were provided with a description of the culprit (M=.18 versus M=.20, in the description and no description conditions, respectively). As shown in Table 1, in 2 out of 11 lineups, mock witnesses in the description condition identified the suspect at a rate above chance expectation (.17). These suspects were not identified above chance expectation, however, in the no description condition. Instead, two other suspects were identified at above chance rates in the no description condition. Additionally, description condition did not systematically influence either functional size (M=8.59 versus M=7.44, in the description and no description conditions, respectively) or Tredoux' E (M=3.87 versus M=4.23, in the description and no description conditions, respectively). Thus, the majority of the lineups were fair based on these traditional measures of lineup bias and size.

Self-Reports of Decision Processes

Mock witness self-reports of decision processes by description condition and the type of reason given are presented in Figure 1. As shown, criminal appearance was most often reported as the primary reason for mock witness choices in the no description condition, representing 32% of all lineup choices made. In the no description condition, the physical appearance of the lineup member having matched the description was reported most often as the primary reason for selection in 45% of all identifications made. Thus, as predicted, when mock witnesses had nothing else on which to base their identification, they were likely to utilise criminal appearance

to make a lineup choice.

The following are representative examples of actual responses given by participants that were coded positive on the criminality criterion:

"Looks like the hicks from my hometown that always got away with beating their wives."
"Fits stereotype of killer on TV"
"His eyes give me the heebee-jeebees"
"Seems mean and evil, like he's up to no good."
"Looks like a drug addict that will do anything to get what he wants."
"I'd switch sides of the street if I saw him walking on my side."

Criminal appearance was reported as the basis for the identification significantly more often in the no description compared to the description condition. In the no description condition, 32% of all identifications were reportedly based on the person appearing criminal, whereas criminality was reported significantly less often in the description condition, representing only 9% of all identifications, $\chi2(1)=45.59$, $p<.01$. In the no description condition, 84% of the mock witnesses reported at least once to have utilised criminality as a criterion, whereas only 52% of mock witnesses reported this was so in the description condition, $\chi2(1)=4.23$, $p<.05$, with Yates correction applied. Examples of participant responses that were coded in the affirmative for guilt included:

"Looks guilty and like he is going to cry."
"Looks nervous."
"Looks like he expects to be picked."
"Looks the most scared."
"Looks suspicious, all drugged up, probably trying to escape the reality of the situation."
"Looks like he is lying."

In the no description condition, 72% of mock witnesses indicated at least once that they used guilt as the criterion for selecting the lineup member (19% of all identifications were reportedly made using guilt). Mock witnesses in the description condition were significantly less likely to

report having used guilt, with only 43% of mock witnesses reporting guilt (6% of all

identifications were reportedly based on guilt), $\chi2(1)=45.59$, p <.01.

The reporting of criminality and guilt seemed to be independent of one another, as mock

witnesses never reported having used both to make their selection. Additionally, the number of

times that a given face was identified based on guilt was unrelated to the number of times that

the face was chosen based on criminal appearance, r=.10, N=66, p=.42. This result suggests that

criminal appearance is a unique dimension of faces that might be separable from guilt.

Other types of reasoning that could be distinguished included: the emotion conveyed by the

lineup member (14% of all identifications made in the no description condition and 4% in the

description condition); the "look" the lineup member had in his eyes (11% of all identifications

in the no description condition, and 2% in the description identifications); the lineup member

was different in physical appearance compared to the other members (1% of all identifications in

the no description condition and 9% in the description condition); and the lineup member was

either smiling or not smiling (7% of all identifications in the no description identifications, and

<1% in the description condition). No other major categories were apparent from the self-report

data. The following are examples of reasons that were not categorized: "gut feeling", "most clean

cut", "I don't know" and "his style of dressing."

Were suspects more often than foil faces chosen based on criminal appearance?

The rate at which a face was identified based on criminal appearance did not differ depending

on whether the face was a suspect or a foil face. In the description condition, 13% of suspect

identifications and 10% of foil identifications were reportedly made based on criminal

appearance. In the no description condition, 30% of suspect identifications and 32% of foil

identifications were reportedly made based on criminal appearance. Thus, based on the selfreport

data, the suspects did not appear to be more criminal looking than the foils.

Discussion

The sample of lineups seemed to be fair based on traditional measures of lineup fairness and suspect bias. Providing mock witnesses with a description of the culprit did not systematically affect measures of lineup size or the probability that the police suspect was identified. Mock witnesses, however, were more likely to report that they had used criminal appearance to make a lineup choice when no description of the culprit was provided. This result suggests that eyewitnesses may use criminal appearance as an alternative identification strategy when they have no information about the culprit's appearance and yet are required to pick someone from a lineup.

Study 1 found that participants were equally likely to use criminal appearance in selecting a foil or the suspect. This might lead one to conclude that the suspect tended to be no more criminal-looking than any of the foils faces were. This conclusion, however, is weakened by the fact that criminal appearance was not systematically measured for all of the lineup faces.

Additionally, participants retrospectively self-reported only a single reason for their identification. Criminal appearance may very well have figured into their choice, but the method that was used to measure criminal appearance was not sensitive enough to capture it. Therefore, the purpose of Study 2 was to systematically measure the criminal appearance of each lineup face to test whether criminal appearance is associated with mock witness identification outcomes. If criminal face bias affects mock witness identifications, then faces that are rated high with respect to criminal appearance should be identified more often than faces rated lower in criminal appearance. Additionally, lineup member similarity and face typicality and distinctiveness ratings were obtained to test whether criminal appearance can predict mock

witness choices once physical appearance is taken into account.

STUDY 2

Method

Participants

Sixty-nine undergraduates (58% female) participated in the face-rating portion of the study; 39 were randomly assigned to one of the rating conditions (criminality, typicality, or distinctiveness, n = 13 in each) and 30 were assigned to rate the physical similarity of the lineup members. The mean age of the raters was 19.86 (SD = 1.50) years. The majority of the participants indicated they were either Caucasian (40%) or Asian (27%). Course credit was awarded for participation.

Materials

Faces (n = 66) were cropped from the lineups employed in Study 1. The faces were each saved as individual image files. Consequently, the faces displayed to raters were identical in all respects to how they appeared in the original lineup.

Design and Procedure

The rating tasks were computer-administered. A program was written in Visual Basic to display the photographs in a random order and to record participant responses. Ratings were made using a scale, which appeared on-screen below the stimulus. Participants clicked on the portion the scale that corresponded with their rating. If desired, re-clicking the scale would change the response. Responses were submitted by pressing the "enter" key. No response deadline was imposed.

Participants individually completed the rating tasks and were randomly assigned to one of them (criminality, distinctiveness, typicality, or similarity). Participants were asked to focus on

the physical appearance of the person in all of the rating tasks. They were told to ignore clothing, focal size, picture size, etc, and to use their own judgment and experience in making the ratings.

In the criminality, distinctiveness, and typicality rating tasks, the presentation order for the 66 individual photographs was randomized for every rater. The photographs were presented one at a time. In the criminality-rating task, participants were asked to rate "the extent to which the face resembled a criminal." For definitional purposes, participants were told that criminal means "someone who would break the law." The rating scale was anchored at 0, "not at all criminal", and 100, "completely criminal." For the distinctiveness-rating task, participants were instructed to rate "the distinctiveness of the face, or the extent to which the face would stand apart from other faces." The rating scale was anchored at 0, "not at all distinctive" and 100, "completely distinctive." For the typicality-rating task, participants were asked to rate the "typicality of the face, or the extent to which the face would resemble other faces." The rating scale was anchored at 0, "not at all typical", and 100, "completely typical."

With respect to the similarity-rating task, two photographs from a given lineup were presented pair wise, or side-by-side, for comparison. There were 15 pair wise evaluations possible for every lineup; therefore, each participant made 165 pair wise evaluations across the 11 lineups. The order in which the pairs were presented was randomized for every subject.

Participants were asked to determine the extent to which the face pairs were "physically similar in appearance to one another." They were instructed to attend to only the physical attributes of the person, including facial features, complexion, hairstyle, and hair colour (dark or light, as the images were black and white) and to ignore other features, such as clothing or the size of the picture. The rating scale was anchored at "0", "not at all similar", and "100", completely similar."

Measures

 Face Rating Data. The criminality, distinctiveness and typicality scores were each averaged across raters for every face. Two-way random intraclass correlation analyses were performed on the rating data for the faces in each lineup; results indicated adequate levels of reliability (criminality: average alpha = .74, range = .61-.86; distinctiveness: average alpha = .83, range = .71-.96; typicality: average alpha = .80, range = .62-.90).

 Similarity Rating Data. The similarity of each lineup member with respect to the other lineup members was measured by averaging within rater each face's pair wise similarity ratings. These data were then averaged across raters and resulted in an average similarity rating for every lineup member. The similarity ratings were deemed reliable given the results of the two-way intraclass correlation analyses (average alpha = .86, range = .82-.94).

Results

Preliminary Results

 Typicality and distinctiveness ratings were significantly associated (r = -0.78, p < .001). The divergent association indicates that participants had rated the members with respect to distinctiveness in the manner intended. Because of the strong association between distinctiveness and typicality that was observed, only distinctiveness was entered in the analyses that follow. Distinctiveness ratings were also significantly associated with criminality ratings (r = .67, p < .01). In subsequent analyses, therefore, both measures were examined in relation to mock witness identification outcomes.

Criminal Appearance and Mock Witness Choices

 The associations between mock witness choice rates and each of the face rating measures (criminality, distinctiveness, and similarity) were examined within each lineup using Pearson's r.

Correlation estimates with small sample sizes are improved by averaging coefficients across samples (Silver & Dunlap, 1987). Therefore, r was calculated and transformed to Fisher's z for every lineup. The obtained z-scores were then averaged across the 11 lineups. The average was backtransformed to r and tested for statistical significance.

In the no description condition, criminal appearance ratings were significantly associated with mock witness identifications (r = .50, z = 2.34, p < .01). The heterogeneity of the relationship between criminal appearance and identification rates across the 11 lineups was tested. The results indicated that the association was consistent across lineups, $\chi2(10) = 5.18$, p = .88. Distinctiveness (r = .25, z = 1.17, p > .05) and similarity (r = -0.05, z = -.23, p > .05) were not associated with identification rates in the no description condition.

In the description condition, the face ratings were not significantly associated with identification outcomes (criminality: r = -0.11, z = -0.52; distinctiveness: r = -0.05, z = -0.23; and similarity: r = 0.25, z = 1.19, all p's > .05).

The results thus far indicate that the criminal appearance ratings postdicted mock witness identifications in the no description condition, whereas the other appearance measures did not. Since criminal appearance is associated with distinctiveness, partial correlation analysis was undertaken to test whether criminal appearance accounted for suspect identifications after face distinctiveness was taken into account. Criminal appearance was indeed significantly associated with mock witness identifications after controlling for distinctiveness, $\rho r = .35$, p < .01, as well as after controlling for lineup member similarity, $\rho r = .40$, p < .01.

Suspect's Criminal Appearance Compared to Other Lineup Members

The next set of analyses examined how often the suspect was the most criminal looking member of the lineup. Table 1 provides suspect identification rates, and the suspect's ranking in

the lineup with respect to criminality, distinctiveness and similarity. Six out of 11 suspects were ranked the highest, either a 1 or 2, with respect to criminality. Moreover, 6 out of 11 suspects were ranked the highest, at either a 1 or 2, with respect to distinctiveness. With regard to similarity, 6 out of 11 suspects were ranked the lowest, at either a 5 or a 6.

In the no description condition, the rate at which the suspect was identified was significantly related to his criminal appearance ranking (Spearman's r = -.56, p < .05). Distinctiveness (Spearman's r = -0.45, p = .08) and similarity (Spearman's r = -0.16, p > .05), however, were not significantly associated with suspect identification rates in the no description condition. None of the rankings were associated with suspect choice rates in the description condition.

Validity of Self-Reported Use of Criminality

The proportion of participants reporting having used criminality as a criterion for a given face was correlated with the face-ratings (criminality, distinctiveness, and similarity) that were made by the independent raters. In the no description condition, the proportion of witnesses indicating they had used criminality as a criterion was significantly related to the independent ratings of criminality (r = .30, p < .05) and distinctiveness (r = .21, p < .05); similarity (r = .05, .p = .68) was not related to self-reports of having used criminality. In the description condition, the proportion of witnesses indicating they had chosen the face because it appeared criminal was unrelated to any of the independent face ratings.

With regard to self-reported use of guilt as a criterion, the proportion of witnesses basing their decision on guilt was negatively associated with distinctiveness ratings (r = -.23, p < .05) in the no description condition; similarity and criminality were not related to guilt. If participants had a description of the culprit, faces that were chosen on the basis of guilt tended to look more similar to the other lineup members (r = .33, p < .01); criminal appearance and distinctiveness were not

related to guilt.

GENERAL DISCUSSION

The results indicated that criminal appearance affected mock witness choices when no description of the culprit was provided. In Study 1, criminal appearance was most often selfreported by mock witnesses as the basis for selecting a face. When a description of the culprit was provided, mock witnesses indicated that their choice was influenced by the description rather by criminal appearance, a result that is in keeping with MacLin et al. (2001). In Study 2, faces that were rated relatively high with respect to criminality were chosen more often in the no description condition; similarity and distinctiveness ratings were unrelated to lineup choices in the absence of a physical description. Ratings of criminality, distinctiveness and similarity were not related to mock witness choices in the description condition.

Since criminal face bias was not associated with identifications when mock witnesses had a suspect description, this raises the possibility that criminal face bias may play a limited role in actual eyewitness identifications. On the one hand, we agree. If the description condition simulates real world circumstances in which eyewitnesses have a memory for the perpetrator, then perhaps criminal face bias will play little to no role in eyewitness identification processes.

On the other hand, there may be real world circumstances in which criminal face bias may play a larger role, such as when eyewitness memory for the perpetrator is weak or when eyewitnesses are highly suggestible, and/or prone to guess. Under these conditions, the operation of criminal face bias may lead to the selection of the suspect, who may or may not be guilty, if he looks more criminal than the other lineup members. Further research is needed to examine the effects of criminal face bias on lineup identifications with vulnerable witnesses.

The results further indicate that criminal appearance should be taken into account along with

the physical appearance in constructing lineups. In more than half of the lineups sampled, the suspect was rated as the most or second most criminal-looking member. All of the foil photographs were drawn from a mug shot database in the current study. Hence, the context in which their photographs were taken was similar—they were all under arrest at the time. In lineups in which the suspect photo is a mug shot and the other lineup members are police officers, public volunteers, or taken from driver license databases, the suspect may especially pop out from the others because his photograph was taken under different circumstances. Put differently, people may be more likely to display the emotions and physical traits that are associated with criminality when they are under arrest. Therefore, it may be the case the criminal face bias effect will vary depending on the database that is used to select foils.

Further research is needed to identify the features that are associated with the criminal stereotype and how they affect lineup decision processes. The specific elements associated with criminal face stereotypes have not yet been identified (MacLin & Herrera, 2006). A ripe area for further research is to discover what aspects of faces give rise to criminal appearance. Research that identifies what these features are could be used to shape lineup construction procedures. The results of the present project suggest that in the meanwhile, assessments of lineup fairness should include not only measures of physical similarity, but also measures of criminal appearance. In particular, the mock witness test conducted without a description could provide information concerning whether the members are matched with respect to criminal appearance.

We further found that criminality appears to be a property of faces unto itself. Though criminal appearance was positively associated with distinctiveness ratings, only criminal appearance was positively associated with mock witness identifications in the no description condition. Lineup member similarity and distinctiveness did not predict mock witness choices

when mock witnesses did not have a description. These results suggest that measuring solely the distinctiveness and/or the physical similarity of the lineup members may be inadequate as a means of determining whether a lineup is biased for an eyewitness who has no memory for the culprit but who is willing to guess. Thus, the criminal appearance of lineup members should perhaps be routinely assessed along with measures of lineup member similarity to determine whether the lineup is fair (i.e., in a fair lineup, choices—in the absence of memory for the culprit—will not be biased either toward or away from the suspect).

We also discovered from participants' self-reports that an appearance of guilt affected their choices. Interestingly, guilt may be a construct that is separable from criminality. Participants most often reported using one or the other to identify faces. It is still possible, however, that guilt and criminality are constructs that overlap to a large extent. Participants were asked to provide the reason(s) behind their decision in an open response format. The results may have been different in a closed format, wherein guilt and criminality were presented as separate response options. Further research is needed to determine the degree to which guilt and criminality overlap in characterizing faces, and how each of these factors affects lineup identifications.

Some of the limitations of the present study should be addressed before concluding. Only 11 lineups could be sampled. Clearly, replication is warranted with randomly selected police lineups, and we hope that researchers and law enforcement can work together in this regard. Previous studies that have assessed the fairness of police lineups from actual criminal cases in the United States have analysed 10 or fewer non-randomly selected lineups (Buckhout et al., 1988; Corey, Malpass, & McQuiston, 1999; Reed, 1984; Wells & Bradfield, 1999); one study analysed 19 non-randomly selected photospreads (Brigham et al., 1999). Studies conducted in the United Kingdom have used larger samples of randomly selected lineups (Valentine &

Heaton, 1999; Valentine, Harris, Colom Piera & Darling, 2003). Sample sizes tend to be small, especially in the U.S., because it is difficult for researchers to obtain lineups from actual cases. U.S. researchers typically obtain lineup stimuli from their consultations with defense attorneys about potentially problematic lineups. In the present study, we were able to randomly select lineups; however, we had great difficulty locating clear photocopies of lineups within the timeframe that we had been allotted to search the police files. Hence, lineup sample size was small. However, our goal in the present study was not to estimate the fairness of police lineups in general. For such an endeavor, additional police jurisdictions and a larger number of lineups would have been needed. Rather, our goals were to 1) examine the association between criminality and mock witness choices, and 2) examine in a preliminary fashion whether there is any evidence that criminality might influence police lineup construction in actual criminal cases. Lineup sample size in the present study was arguably adequate for these purposes.

Another potential limitation of the current study's findings is that black and white rather than colour photographs were utilised as lineup stimuli. The criminal face bias effect may have been larger had colour photographs been used, as colour photographs may better convey emotional states and fine features of faces associated with criminality, such as pock marks and scars (MacLin & Herrera, 2004). The type of crime committed may also influence the size of the criminal face bias effect. We did not inform participants of the type of crime the person was suspected of committing. The use of criminal face stereotypes may have been more prevalent if we told participants that the crime was murder or rape, and less prevalent if the crime was forgery. Lastly, little is known about female criminal stereotypes (see Bustamante, Herrera, & MacLin, 2001). Perhaps stereotyped views of female compared to male criminality would play a larger role in lineup identification. For instance, because of their relative rarity as murderers,

stereotypes of women who commit murder may be especially polarized (see Zebrowitz-McArthur, 1982 for a theoretical overview of polarization in the development of stereotypes in general). Therefore, criminal face bias may possibly operate to a larger extent when the perpetrator is a woman.

In sum, criminal face bias appears to be fertile ground for continued research. This research could identify which features of faces are associated with criminal stereotypes and subsequently cause some faces to stand apart from other faces. Criminal face bias may play a role in criminal lineup identifications at the level of the eyewitness and at the level of lineup construction. Eyewitnesses who hold stereotyped views about the appearance of criminal perpetrators may base their lineup identifications on this information. Moreover, lineups may be constructed, either consciously or inadvertently, such that the suspect is more criminal in appearance than the other members, which in turn may increase the possibility that the suspect stands apart from the foils. The results from the present project indicates that if criminal face bias is allowed to act as an extra-memory influence in a real world lineup, a lineup identification opportunity may be thwarted.

References

Brigham, J. C., Meissner, C. A., & Wasserman, A. W. (1999). Applied issues in the construction and expert assessment of photo lineups. Applied Cognitive Psychology, 13, S73-S92.

Buckhout, R., Rabinowitz, M., Alfonso, V., & Kanellis, D. (1988). Empirical assessment of lineups: Getting down to cases. Law and Human Behavior, 12, 323-331.

Bull, R. (1979). The influence of stereotypes on person identification. In D. P. Farrington, K. Hawkins, & S. M. Lloyd-Bostock (Eds.), Psychology, Law, and Legal Processes (pp. 184-194). London: Macmillan Press.

Bull, R. H. C., & Rumsey, N. (1988). The social psychology of facial appearance. New York: Springer.

Bustamante, G., Herrera, V., & MacLin, M. K. (April, 2001). The masculinity-femininity link: Effects on mock witness identification of female targets. Poster presented at the Rocky Mountain Psychological Association Convention, Reno, NV.

Corey, D., Malpass, R., & McQuiston, D. (1999). Parallelism in eyewitness and mock witness identification. Applied Cognitive Psychology, 13, S41-SS58.

Doob, A. N. & Kirshenbaum, H. M. (1973). Bias in police lineups--Partial remembering. Journal of Police Science and Administration, I(3), 287-293.

Flowe, H. D., Ebbesen, E. B., Libuser, M., Rienick, C., & Van Ness N. (2010). Testing the Reflection Assumption: A Comparison of Eyewitness Ecology in the Laboratory and the Field. Manuscript under review.

Goldstein, A. G., Chance, J. E., & Gilbert, B. (1984). Facial stereotypes of good guys and bad guys: a replication and extension. Bulletin of the Psychonomic Society, 22, 549-552.

Lombroso, C. (2005). Criminal Man. Durham, NC: Duke University Press. (Original work
published 1911.)

MacLin, M. K., & Herrera, V. (2006). The Criminal Stereotype. North American Journal of
Psychology 8(2):197–208.

MacLin, O. H. & MacLin, M. K. (2004). The Effect of Criminality on Face Attractiveness,
Typicality, Memorability and Recognition. North American Journal of Psychology, 6(1),
145- 154.

MacLin, M. K., Malpass, R. S., & Herrera, V. T. (2001). The effect of criminality on
mock witness identification. Paper Presented at the Annual Meeting of the Society for Applied
Research in Memory and Cognition, Kingston, Ontario, Canada.

MacLin, M. K., Waack, B., & Peterson, H. (2003). Criminal appearance stereotypes: Personality
traits, facial features and crime type. Legal Psychology Program and International
Forensic Research Institute at Florida International University Symposium, Miami, FL.

Mocan, H. N., & Tekin, E. (2005). Ugly Criminals. Retrieved from
http://www.aeaweb.org/annual_mtg_papers/2006/0106_0800_0902.pdf

Reed, T. E. (1984). Eyewitness identifications of an armed robber within a biased police lineup.
Journal of Police Science Administration, 12, 310-314.

Saladin, M., Saper, Z., & Breen, L. (1988). Perceived attractiveness and attributions of
criminality: What is beautiful is not criminal. Canadian Journal of Criminology, 30, 251-
259.

Shoemaker, D. J., South, D. R., & Lowe, J. (1973). Facial stereotypes of deviants and
judgments of guilt or innocence. Social Forces, 51, 427–433.

Silver, N., & Dunlap, W. (1987). Averaging correlation coefficients: Should Fisher's z transformation be used? Journal of Applied Psychology, 72(1), 146-148.

Tredoux, C. G. (1998). Statistical inference on lineup measures. Law and Human Behavior, 22(2), 217-237.

Valentine, T., Harris, N., Colom Piera, A., & Darling, S. (2003). Are police video identifications fair to African-Caribbean suspects? Applied Cognitive Psychology, 17, 459-476.

Valentine, T., & Heaton, P. (1999). An evaluation of the fairness of police line-ups and video identifications. Applied Cognitive Psychology, 13, S59-SS72.

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. and Ostrom, T. M. (1979). Guidelines for empirically assessing the fairness of a lineup. Law and Human Behavior, 3, 285-293.

Yarmey, D. A. (1993). Stereotypes and recognition memory for faces and voiced of good guys and bad guys. Applied Cognitive Psychology, 7, 419-431.

Zebrowitz-McArthur, L. (1982). Judging a book by its cover: A cognitive analysis of the relationship between physical appearance and stereotyping. In A. H. Hastorf & A. M. Isen (Eds.), Cognitive social psychology (pp. 149–209). New York: Elsevier.

Footnote

1 Asian refers to Chinese, Korean, Japanese, or Filipino ancestry in the geographical region in

which the study was conducted.

Table 1.

Suspect identification rates, rankings for criminality, distinctiveness and similarity, and the

lineup fairness data for the randomly selected police lineups.

Lineup	Suspect Position	Mock Witness Suspect		Suspect Rankings				Lineup Fairness Description Condition			Lineup Fairness No Description Condition		
		Desc (n = 23)	No Desc (n = 25)	Criminal	Distinctive	Typical	Similar	Bias	Functional Size	Tredoux' E	Bias	Functional Size	Tredoux' E
1	5	0.39*	0.04	5	6	2	5	Yes	2.56	3.81	No	25.00	3.93
2	3	0.17	0.48**	1	1	6	4	No	5.75	1.92	Yes	2.08	3.02
3	3	0.17	0.36*	1	1	5	4	No	5.75	5.45	Yes	2.77	4.08
4	3	0.43*	0.12	2	2	2	1	Yes	2.20	3.46	No	8.33	3.49
5	2	0.04	0.24	2	1	4	1	No	23.00	4.52	No	4.17	4.37
6	3	0.00	0.20	1	3	3	6	No	undef	2.10	No	5.00	3.93
7	5	0.21	0.16	3	4	6	5	No	4.60	5.14	No	6.25	5.53
8	3	0.04	0.16	6	1	5	6	No	23.00	2.63	No	6.25	4.70
9	2	0.13	0.16	6	5	1	2	No	7.67	4.85	No	6.25	5.08
10	3	0.13	0.12	2	1	6	5	No	7.67	3.98	No	8.33	4.63
11	3	0.26	0.20	4	5	4	5	No	3.67	4.75	No	5.00	3.83

*p < .05. **p < .01. †Functional size is undefined because the suspect was never chosen by mock witnesses.
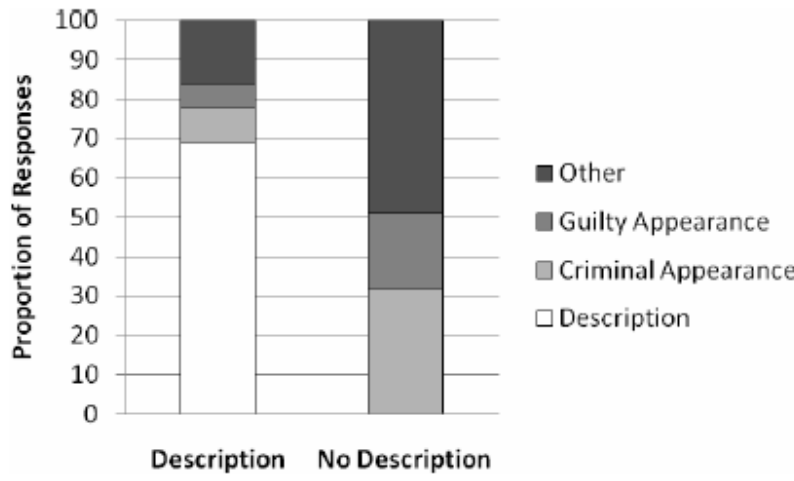
Figure 1

Figure 1. Distribution of self-reported reasons for having selected a given face within the description and no description conditions in Study 1.