

Selecting Fillers on Emotional Appearance Improves Lineup Identification Accuracy

Heather D. Flowe, Thimna Klatt, and Melissa F. Colloff
University of Leicester

Mock witnesses sometimes report using criminal stereotypes to identify a face from a lineup, a tendency known as *criminal face bias*. Faces are perceived as criminal-looking if they appear angry. We tested whether matching the emotional appearance of the fillers to an angry suspect can reduce criminal face bias. In Study 1, mock witnesses ($n = 226$) viewed lineups in which the suspect had an angry, happy, or neutral expression, and we varied whether the fillers matched the expression. An additional group of participants ($n = 59$) rated the faces on criminal and emotional appearance. As predicted, mock witnesses tended to identify suspects who appeared angrier and more criminal-looking than the fillers. This tendency was reduced when the lineup fillers matched the emotional appearance of the suspect. Study 2 extended the results, testing whether the emotional appearance of the suspect and fillers affects recognition memory. Participants ($n = 1,983$) studied faces and took a lineup test in which the emotional appearance of the target and fillers was varied between subjects. Discrimination accuracy was enhanced when the fillers matched an angry target's emotional appearance. We conclude that lineup member emotional appearance plays a critical role in the psychology of lineup identification. The fillers should match an angry suspect's emotional appearance to improve lineup identification accuracy.

Keywords: criminal face bias, lineup bias, eyewitness identification, simultaneous lineup

People have well-formed stereotypes about the physical traits of criminals. Stereotypical criminal perpetrators are deemed to be unattractive (Saladin, Saper, & Breen, 1988), have long or shaggy dark hair, tattoos, beady eyes, pock marks, and scars (MacLin & Herrera, 2006). People also readily agree on whether a face looks criminal (Bull & Green, 1980; Bull & McAlpine, 1998; Goldstein, Chance, & Gilbert, 1984; Macrae & Shepherd, 1989; Shoemaker, South, & Lowe, 1973; Valla, Ceci, & Williams, 2011; Yarmey, 1993). Although research has shown that facial appearance has little to no validity for predicting actual behavior (Olivola & Todorov, 2010), stereotypes about criminal appearance have been found to affect legal decision making (e.g., Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006; MacLin, Downs, MacLin, & Caspers, 2009; Zebrowitz & McDonald, 1991; also see Porter & ten Brinke, 2009). Here, we examined a method for reducing the effects of criminal face bias on eyewitness identifications.

An eyewitness' identification should be based on memory. Yet, criminal stereotypes may influence the outcome of a lineup identification, whereby witnesses select the most criminal-looking person from the lineup. Hereafter, we will refer to this tendency as *criminal face bias*. To illustrate, Wells and Bradfield (1999) found that mock witnesses identified the police suspect more often when

asked, "Which is the accused?" compared to when they were asked to select the face in the lineup that matched the eyewitness' description. Mock witnesses are given a lineup test, but they do not view a crime or a to-be-remembered face (Doob & Kirshenbaum, 1973). If mock witnesses identify the suspect at greater than chance expectation without having actually seen the perpetrator, this indicates the suspect is distinct from the fillers and we can conclude that the lineup is biased (Malpass & Lindsay, 1999). Thus, Wells and Bradfield's results suggest that mock witness choices can be biased by people's expectations about what an accused person looks like.

Other work further suggests that mock witness identifications can be influenced by criminal face bias (Flowe & Humphries, 2011; McQuiston & Malpass, 2002). Flowe and Humphries (2011) obtained a random sample of photographic lineups that had been used in actual criminal investigations and presented them to mock witnesses. The police had composed the lineups using mugshots. Half of the mock witnesses received a description of the culprit and the other half did not. Another group of people rated how criminal-looking each of the lineup faces appeared. Mock witnesses who did not receive a description of the culprit were biased toward selecting the faces that were rated as the most criminal-looking. When these participants were asked an open-ended question regarding why they had chosen a particular face, they stated that it was because the person looked like a criminal. Additionally, approximately half of the suspects were rated as more criminal-looking than the fillers and they were identified from lineups by mock witnesses at a level greater than chance. The researchers concluded that criminal appearance has the potential to bias lineup identifications, particularly when memory for the culprit is weak or nonexistent and the witness feels compelled to identify someone.

This article was published Online First August 18, 2014.

Heather D. Flowe, Thimna Klatt, and Melissa F. Colloff, School of Psychology, University of Leicester.

Thimna Klatt is now at Christian-Albrechts-University Kiel. Melissa F. Colloff is now at University of Warwick.

Correspondence concerning this article should be addressed to Heather D. Flowe, University of Leicester, School of Psychology, Forensic Section, 106 New Walk, Leicester, UK, LE1 7EA. E-mail: hf49@le.ac.uk

Guidelines for the creation of fair lineups advise that the suspect should not stand out in the lineup (Technical Working Group for Eyewitness Evidence, 1999; Wells et al., 1998). To date, however, research has not examined how to prevent criminal-looking suspects from standing apart from the fillers. In the present study, we tested whether criminal face bias can be reduced by matching the fillers to the suspect on emotional appearance, and in particular, on anger. Faces are rated as less trustworthy when they display anger or have morphological features that connote anger (e.g., Neth & Martinez, 2009; Oosterhof & Todorov, 2008; Said, Sebe, & Todorov, 2009). Angry faces are also evaluated as criminal-looking (Flowe, 2012). One interesting question that arises from this research is whether criminal face bias can be reduced if the fillers match the emotional appearance of an angry-looking suspect.

In Study 1, we examined mock witness identification rates across lineups in which the emotional appearance of the fillers relative to the “suspect” was varied. We operationally defined emotional appearance as the overt emotional expression displayed on the face. The suspect’s emotional expression was angry, happy, or emotionally neutral, and the fillers’ emotional expressions either matched the suspect’s or were emotionally neutral. Additionally, a group of raters evaluated the faces to establish whether criminal appearance was associated with perceptions of anger.

We hypothesized that mock witnesses would identify a suspect with an angry emotional expression more often when he or she was surrounded by neutral as opposed to angry fillers. Additionally, mock witnesses were expected to select a happy suspect less often when he or she was surrounded by neutral rather than happy fillers. Previous research indicates that a neutral face should appear relatively angrier and more criminal-looking compared to a happy face, all other things being equal (Flowe, 2012). Hence, mock witnesses will direct their choice to a neutral rather than happy face if they are selecting the face that appears relatively the most criminal. We also examined the effective size of the lineups in relation to suspect and filler emotional appearance. Effective size captures how many members in the lineup are plausible alternatives to the suspect (Malpass, 1981; Wells, Leippe, & Ostrom, 1979). Alternatives are plausible to the extent that they are similar in appearance to the suspect. We hypothesized that if emotional appearance affects mock witness identifications, then the effective size will be larger in emotion-matched compared to unmatched lineups. Additionally, if mock witnesses are predisposed to select angry-looking faces because they appear criminal, then the effective size of the lineup will be larger when all of the members have an angry expression as opposed to a different type of emotional expression. Finally, we tested our hypotheses with both male and female suspects. Although lineup research rarely includes female faces, there is some evidence suggesting that the relationship between angry expressions and criminal appearance should hold for both male and female faces (Flowe, 2012). Both male and female faces are evaluated as more criminal-looking if their faces appear angry. Given this, we did not expect the results to vary by suspect gender.

Study 2 examined whether the findings obtained with mock witnesses extended to recognition memory. Participants’ memory for a previously studied male face was tested using a simultaneous lineup. We varied how angry the suspect appeared and whether the fillers matched the suspect’s emotional appearance.

Study 1

Method

Design. A 3 (suspect emotional expression: angry vs. happy vs. emotionally neutral) \times 2 (filler type: matched vs. unmatched) \times 2 (suspect gender: man vs. woman) factorial design was employed. The outcome measure was the suspect identification rate.

Participants. A group of participants ($n = 59$) rated face attributes and another group ($n = 226$) participated as mock witnesses. Participants ranged in age from 18 to 65 ($M = 30.84$ years; 40% female) and were recruited from the University of Leicester and the general Internet population via Amazon Mechanical Turk. In exchange for participation, the University participants were given course credit and the Mechanical Turk participants were paid 20 cents (USD). The purpose of including the Internet sample was to increase the age range and number of male participants. The majority of participants were Caucasian (30%) or Asian (61%).

Materials. The suspects and fillers were drawn from the Radboud Faces Database (Langner et al., 2010), which contains photos of adult males ($n = 20$) and females ($n = 19$), aged 21 years on average. The database has a photograph of each person displaying different emotional expressions using the Facial Action Coding System (FACS) (Ekman & Friesen, 1976). Photographs of the angry, happy, and emotionally neutral facial displays were utilized. We composed lineups with faces from this database because the emotion displays have been validated. For example, 85% and 96% of participants in the validation study chose the target emotion for angry and happy displays, respectively (Langner et al., 2010). Additionally, the technical elements of the photographs (e.g., lighting, focal distance), clothing, and age are held constant across the set of photos, and the individuals do not have distinctive features (e.g., scars, tattoos, earrings). However, the Radboud faces do look dissimilar to each other on some physical features (e.g., hair color, eye color). Hence, for a lineup in which the suspect appears angry but the fillers are emotionally neutral, the suspect and fillers will differ on other dimensions besides emotional appearance. One might wonder whether we should have used lineups in which the individuals were more physically similar to each other. If anything, however, the fact that we had variation across lineup members in physical appearance should have made it more difficult for us to find support for our hypotheses. Imagine we had composed lineups with faces that were very similar in physical appearance. The emotional expression of the faces would have been a highly salient feature under these circumstances. With limited variation in appearance across faces, it would not have been surprising if emotional expression impacted mock witness identifications because it essentially would be one of the few features that could be used to distinguish one face from another. Therefore, we employed the Radboud faces to avoid this criticism and to provide a more conservative test of our hypothesis.

For each person in the database, we created a simultaneous lineup in which he or she was the “suspect.” In total, there were 19 female lineups and 20 male lineups. Each lineup contained five same-gender fillers and was displayed in a 2×3 array. The fillers were randomly drawn from the other same-gender faces in the database. Note that in order to create the number of lineups

needed, each face served as a filler between five and six times across the lineups for a given gender. When a face was randomly selected as a filler six times, it was not used as a filler in any further lineups. For each lineup version (see below), we correlated the number of times that each face was repeated across the set of lineups with how often the face was identified when it was the designated suspect. None of the results was statistically significant, indicating that the number of times a face was repeated did not affect how often it was identified (mean $r = .003$, r range: $-.16-.20$).

For each suspect, there were five versions of his or her lineup. The choice of fillers and the position of the lineup members were held constant across the five versions. In three versions, the emotional display of the suspect and fillers matched (i.e., the members were all angry, all happy, or all emotionally neutral), and in two versions, the emotional display of the suspect and fillers was unmatched (i.e., the suspect was angry and the fillers were emotionally neutral, or the suspect was happy and the fillers were emotionally neutral). For each suspect, lineup position number (2, 3, 4 or 5) was randomly determined.

Face rating procedure. The raters were shown individual faces and they evaluated how angry, happy, and criminal-looking the face appeared. We randomly assigned raters to evaluate either all male or all female faces. Male and female faces were rated separately for several reasons. First, lineups are composed of a single gender. Second, female faces tend to look happier than male faces, and male faces appear angrier than female faces (Becker, Kenrick, Neuberg, Blackwell, & Smith, 2007). Consequently, a female face may look happier if rated alongside male faces rather than other female faces (also see Grayson, 1998). As such, ratings made to a female face when she is evaluated alongside male faces would potentially be less predictive of how people would perceive her face when she is in a lineup with other females. Based on a similar logic, we also randomly assigned participants to rate all angry, all happy, or all emotionally neutral faces. This approach should reduce any possible carryover effects whereby, say, a neutral face appears more or less angry depending on whether participants just rated a happy or an angry face. The traits were rated on a 7-point Likert scale, ranging from *not at all [attribute]* to *very [attribute]*.

Mock witness procedure. Mock witnesses were randomly assigned to a lineup condition and they evaluated either all male or all female lineups. Before each lineup was shown, participants were asked, "A police lineup parade appears below. Who do you think is the suspect?"

Analyses. We calculated Cronbach's alpha to assess the reliability of the trait ratings. Cronbach's alpha coefficients for the traits (Males: angry $\alpha = .97$, happy $\alpha = .99$, criminal $\alpha = .91$; Females: angry $\alpha = .97$, happy $\alpha = .99$, criminal $\alpha = .86$) were at an acceptable level of reliability (Kline, 2000). We averaged across raters the anger, happiness, and criminal appearance ratings for each face, conditioning the data on the overt emotion displayed (angry, happy, or emotionally neutral).

The average rate at which mock witnesses identified the suspect and E' (Tredoux, 1998) were calculated for each suspect, conditioning the data on the experimental factors. The suspect identification rate measures lineup bias, or the extent to which choices are directed toward the suspect, whereas E' measures lineup effective size (Malpass & Lindsay, 1999; Malpass, 1981).

Data were analyzed using the general linear model, including t tests and Pearson's r . Alpha was set to .05 in all analyses and all tests were two-tailed. Cohen's d and 95% confidence intervals are reported.

Results

Preliminary results. Table 1 presents the correlations across the anger, happiness, and criminal appearance ratings for male and female suspects by emotional pose. In accordance with previous research, faces rated as more criminal-looking were rated as angrier and less happy-looking. This relationship held for faces that were displaying angry, happy, and emotionally neutral expressions.

Table 2 presents the mock witness identification rates by the experimental factors. We began our analysis by confirming that mock witnesses were not biased toward picking the suspect in matched lineups. If a lineup is not biased against the suspect, the suspect should not be identified at a rate that exceeds chance expectation (i.e., $p = .167$ for a six-person lineup). Indeed, one-sample t tests confirmed that the identification rate for both male and female suspects in matched lineups did not significantly vary from chance expectation in any of the emotional expression conditions (see the top panel of Table 2). We also verified that the suspects did not appear angrier or more criminal looking than the fillers: There was no correlation between the suspect identification rate and the appearance ratings for the matched lineups in which all of the members were making an angry or a happy display (all $ps > .27$). However, for the lineups in which the suspect and the fillers were all making an emotionally neutral display, the suspect identification rate was positively correlated with the anger ($r = .63$, $p < .001$) and the criminal appearance ($r = .60$, $p < .001$) ratings. Thus, even when the suspect and fillers were attempting to appear emotionally neutral, faces that naturally appeared relatively angrier and more criminal-looking were identified more often.

Mock witness identifications of angry suspects. We hypothesized that mock witnesses would identify a suspect with an angry emotional expression more often when he or she was surrounded by neutral as opposed to angry fillers. In keeping with this, an

Table 1
Correlations for Ratings of How Angry, Happy and Criminal the Face Appeared for Male and Female Faces by Suspect Emotional Expression Condition (Neutral, Angry or Happy), Study 1

	Male faces		Female faces		
	Criminal	Angry	Criminal	Angry	
Neutral Faces			Neutral Faces		
Angry	.77**		Angry	.78**	
Happy	-.68**	-.81**	Happy	-.63**	-.49**
Angry Faces			Angry Faces		
Angry	.87**		Angry	.58**	
Happy	-.60*	-.72**	Happy	-.54*	-.59**
Happy Faces			Happy Faces		
Angry	.61**		Angry	.39	
Happy	-.59**	-.66**	Happy	-.55**	-.56**

* Correlation is significant at the .05 level (2-tailed). ** Correlation is significant at the .01 level (2-tailed).

Table 2
Target Identification Rate (Mean, SD and 95% CI for the Mean) by Suspect Emotion, Suspect Gender, and Whether the Emotional Expression of the Suspect and Fillers Matched, Study 1

Suspect emotion	Suspect gender	Mean	SD	CI .95 Lower	CI .95 Upper
Emotion-matched lineups					
Angry	Woman	0.20	0.08	0.16	0.23
	Man	0.17	0.09	0.13	0.21
Happy	Woman	0.15	0.13	0.09	0.21
	Man	0.13	0.09	0.09	0.18
Neutral	Woman	0.14	0.13	0.08	0.20
	Man	0.16	0.13	0.08	0.23
Emotion-unmatched lineups					
Angry	Woman	0.31	0.07	0.28	0.34
	Man	0.22	0.08	0.18	0.26
Happy	Woman	0.05	0.07	0.01	0.08
	Man	0.14	0.11	0.08	0.19

angry suspect surrounded by emotionally neutral fillers was identified at a rate greater than chance expectation ($M = 0.31$, $SD = 0.07$), $t(38) = 6.80$, $p < .001$, $d = 1.08$, 95% CI [0.69, 1.48]. This held for male suspects, $t(19) = 2.88$, $p < .05$, $d = 0.64$, 95% CI [0.15, 1.12] and female suspects, $t(18) = 8.60$, $p < .001$, $d = 1.97$, 95% CI [1.18, 2.74]. Additionally, an angry suspect was identified significantly more often when surrounded by neutral fillers ($M = .26$, $SD = .09$) as opposed to angry fillers ($M = .18$, $SD = .09$), $t(39) = 4.12$, $p < .001$, $d = .65$, 95% CI [.31, .99]. This held for male suspects, $t(19) = 2.01$, $p < .05$, $d = 0.45$, 95% CI [-0.01, 0.90] and female suspects, $t(18) = 3.88$, $p < .001$, $d = 0.89$, 95% CI [0.35, 1.41]. Thus, lineups were biased against an angry suspect if the fillers were neutral as opposed to angry.

The face rating data further supported our hypothesis. Matched samples t tests indicated that angry faces were perceived as significantly angrier, $t(38) = 9.01$, $p < .001$, $d = 1.44$, 95% CI [0.99, 1.89] and more criminal-looking ($t(38) = 2.27$, $p < .05$, $d = 0.36$, 95% CI [0.04, 0.68]) than neutral faces; descriptive statistics for this analysis are shown in Table 3.

Mock witness identifications of happy suspects. We also hypothesized that mock witnesses would be disinclined to pick a happy suspect when he or she was surrounded by neutral fillers, because neutral fillers would appear relatively angrier and more criminal-looking. Indeed, the identification rate for happy suspects was significantly below chance expectation when the fillers were neutral ($M = 0.09$, $SD = 0.10$), $t(38) = -4.57$, $p < .001$, $d = 0.73$, 95% CI [0.37, 1.08]. Additionally, a happy suspect was chosen significantly less often when surrounded by neutral rather than happy fillers, $t(39) = -2.16$, $p < .05$, $d = 0.34$, 95% CI [0.02, 0.66]. The face rating data lent further support for our hypothesis. Matched samples t tests indicated that neutral faces were rated as significantly angrier, $t(38) = 10.82$, $p < .01$, $d = 1.73$, 95% CI [1.23, 2.23] and more criminal-looking ($t(38) = 6.25$, $p < .01$, $d = 1.00$, 95% CI [0.61, 1.38]) than happy faces; descriptive statistics for this analysis are shown in Table 3.

The results for happy suspects, however, were qualified by suspect gender. For happy female suspects, the suspect identification rate ($M = 0.05$, $SD = 0.07$) was significantly below chance

Table 3
Ratings (Mean, SD and 95% CI for the Mean) for How Angry, Criminal, and Happy the Suspect's Face Appeared by Suspect Emotion and Suspect Gender, Study 1

	Angry display			Neutral display			Happy display		
	Angry	Criminal	Happy	Angry	Criminal	Happy	Angry	Criminal	Happy
WOMAN									
Mean	4.48	3.76	1.98	3.47	3.73	2.86	2.35	2.98	5.18
SD	0.73	0.35	0.30	0.51	0.48	0.49	0.25	0.28	0.37
CI .95 Lower	4.16	3.56	1.85	3.21	3.44	2.62	2.22	2.80	5.01
CI .95 Upper	4.81	3.96	2.11	3.73	4.02	3.10	2.47	3.17	5.35
MAN									
Mean	4.73	4.20	2.04	3.32	3.76	3.28	2.47	3.32	4.88
SD	0.68	0.49	0.24	0.59	0.74	0.53	0.29	0.49	0.34
CI .95 Lower	4.41	4.00	1.91	3.07	3.47	3.04	2.34	3.14	4.72
CI .95 Upper	5.05	4.39	2.16	3.57	4.04	3.51	2.59	3.50	5.04
OVERALL									
Mean	4.61	3.98	2.01	3.39	3.75	3.08	2.41	3.16	5.03
SD	0.71	0.48	0.27	0.55	0.62	0.55	0.28	0.43	0.38
CI .95 Lower	4.38	3.84	1.92	3.21	3.54	2.90	2.32	3.02	4.92
CI .95 Upper	4.84	4.12	2.10	3.58	3.94	3.24	2.50	3.28	5.14

expectation when the fillers were emotionally neutral, $t(18) = -7.30, p < .001, d = 1.67, 95\% \text{ CI } [0.96, 2.37]$. Happy female suspects were also identified significantly less often when they were surrounded by emotionally neutral ($M = 0.05, SD = 0.07$) rather than happy ($M = 0.15, SD = 0.13$) fillers, $t(18) = 2.95, p < .01, d = 0.68, 95\% \text{ CI } [0.17, 1.17]$. For happy male suspects, however, the suspect identification rate did not significantly differ from chance expectation when the fillers were emotionally neutral ($M = 0.14, SD = 0.11$), $t(38) = -1.29, p = .21, d = 0.21, 95\% \text{ CI } [-0.11, 0.52]$, nor was the suspect identification rate lower when the suspect was surrounded by neutral rather than happy fillers ($M = 0.14, SD = 0.11$ vs. $M = 0.14, SD = 0.09$, respectively), $t(19) = -0.03, p = .97, d = 0.01, 95\% \text{ CI } [-0.27, 0.28]$. Hence, participants seemed to be indifferent to picking a happy male suspect versus an emotionally neutral male filler.

Effective size in emotion-matched and unmatched lineups.

Our final set of analyses examined whether the effective size of the lineup varied in relation to the experimental factors. E' for matched lineups ($M = 4.24, SD = 0.87$) and unmatched lineups ($M = 4.20, SD = 0.74$) was comparable in size to other research, suggesting the lineups were adequate in size overall (e.g., Gronlund, Carlson, Dailey, & Goodsell, 2009). Table 4 provides descriptive statistics for E' as a function of the experimental factors. For male suspects, E' was significantly larger for matched compared to unmatched lineups when the suspect was happy, $t(19) = 2.53, p < .05, d = 0.57, 95\% \text{ CI } [0.09, 1.03]$ and when the suspect was angry ($M = 5.03, SD = 0.63$ vs. $M = 4.36, SD = 0.73$, respectively), $t(19) = 3.10, p < .01, d = 0.69, 95\% \text{ CI } [0.20, 1.18]$. Contrastingly, in female suspects, E' was significantly larger for matched compared to unmatched lineups when the suspect was happy ($M = 4.20, SD = 0.76$ vs. $M = 3.72, SD = 0.67$, respectively), $t(18) = 2.06, p < .05, d = 0.47, 95\% \text{ CI } [-0.01, 0.94]$, but did not significantly differ when the suspect was angry ($M = 4.59, SD = 0.55$ vs. $M = 4.65, SD = 0.55$, respectively). Thus, except for angry female suspects, matching the fillers to the emotional appearance of the suspect resulted in lineups that had a larger effective size.

We also hypothesized that the effective size would be larger when all of the lineup members had angry expressions, compared to when all of the lineup members had happy or emotionally

neutral expressions. Matched samples t tests confirmed that E' was significantly larger when all of the members were angry, compared to when all of the members were happy, $t(38) = 3.45, p < .01, d = 0.55, 95\% \text{ CI } [0.22, 0.89]$, or compared to when all of the members were emotionally neutral, $t(38) = 8.17, p < .001, d = 1.30, 95\% \text{ CI } [0.87, 1.73]$. The pattern did not vary by suspect gender. When the suspect was a man, E' was significantly larger when all of the members were angry, compared to when all of the members were happy, $t(19) = 2.62, p < .05, d = 0.58, 95\% \text{ CI } [0.10, 1.06]$, or compared to when all of the members were emotionally neutral, $t(19) = 8.07, p < .001, d = 1.80, 95\% \text{ CI } [1.08, 2.51]$. When the suspect was a woman, E' was significantly larger when all of the members were angry, compared to when all of the members were happy, $t(18) = 2.19, p < .05, d = 0.50, 95\% \text{ CI } [0.02, 0.97]$, or compared to when all of the members were emotionally neutral, $t(18) = 4.36, p < .01, d = 1.00, 95\% \text{ CI } [0.44, 1.55]$.

Discussion

We tested whether criminal face bias can be reduced by matching the fillers to the emotional expression of an angry suspect. In keeping with our hypothesis, mock witnesses identified an angry suspect more often when the fillers were emotionally neutral rather than angry. The emotional appearance of the fillers and suspect also impacted the effective number of fillers the lineups contained, which was measured by E' (Tredoux, 1998). As we had hypothesized, E' was larger when all of the faces had angry expressions compared to when all of the lineup members had either happy or emotionally neutral expressions. Thus, mock witnesses selected angry and criminal-looking faces when they had no other information on which to base their identification.

As we hypothesized, the effective size of a lineup was larger for emotion-matched compared to unmatched lineups, with one exception: Lineup size for angry female suspects did not differ depending on whether the fillers were matched to her emotional expression. Perhaps E' was not smaller for angry female suspects when the fillers were neutral as opposed to angry because some of the neutral female fillers appeared criminal-looking, and hence, attracted mock witness choices. For example, women tend to be rated more negatively if their face is shown in a neutral pose

Table 4
E' (Mean, SD and 95% CI) by Suspect Emotion, Suspect Gender, and Whether the Emotional Expression of the Suspect and Fillers Matched, Study 1

Suspect emotion	Suspect gender	Mean	SD	CI .95 Lower	CI .95 Upper
Emotion-matched lineups					
Angry	Woman	4.59	0.55	4.33	4.86
	Man	5.03	0.63	4.74	5.33
Happy	Woman	4.20	0.76	3.83	4.57
	Man	4.54	0.58	4.27	4.82
Neutral	Woman	3.73	0.68	3.40	4.06
	Man	3.34	0.75	2.99	3.69
Emotion-unmatched lineups					
Angry	Woman	4.66	0.55	4.40	4.92
	Man	4.37	0.73	4.03	4.71
Happy	Woman	3.72	0.77	3.40	4.04
	Man	4.04	0.66	3.72	4.35

(Deutsch, LeBaron, & Fryer, 1987). Even though lineup effective size did not vary for an angry female suspect surrounded by neutral as opposed to angry fillers, it is still important to emphasize that mock witnesses overall were still more biased toward selecting an angry female suspect if the fillers did not match her emotional appearance. Thus, our data support the position that fillers should be matched to the emotional appearance of angry male and female suspects in order to increase the effective size of a lineup.

We also hypothesized that mock witnesses would be biased away from picking a happy suspect when he or she was surrounded by neutral fillers. Suspect identification rates were indeed reduced in lineups in which the fillers looked more criminal than the suspect. However, the results for happy suspects were qualified by suspect gender. Happy female suspects were identified less often when they were surrounded by emotionally neutral rather than happy fillers. For happy male suspects, however, the identification rate did not differ from chance expectation when the suspect was surrounded by neutral fillers, nor was a happy male suspect picked less often when the fillers were neutral as opposed to happy. Perhaps mock witnesses were not less likely to pick a happy suspect over a neutral filler because some happy male suspects still appeared criminal-looking, and hence, attracted choices.

Study 1 has some methodological strengths of note. First, the individuals serving as suspects and fillers were held constant across the experimental conditions. The only factor that varied across conditions was the emotional expression of the faces, which allowed us to conclude that the results were driven by differences in emotional appearance across the fillers and the suspect. Second, the emotional expression of the faces was validated (Langner et al., 2010). Third, it is unlikely that our results were caused by experimenter demand. Namely, one might argue that mock witnesses were simply selecting the face that was visually distinct from the other faces. If true, mock witnesses would have identified the happy suspect and the angry suspect more often in unmatched compared to matched lineups. This result, however, was obtained for only angry suspects, indicating that participants were not simply picking a face because it had a different emotional expression than the other faces.

One outstanding question we had was whether matching the fillers to the suspect on emotional appearance affects recognition memory accuracy. Flowe and Humphries (2011) found that criminal appearance did not predict mock witness identifications when participants had a description of the culprit. Under these conditions, identifications were predicted by how similar the suspect was to the culprit's description. Therefore, criminal face bias did not occur when participants had other information on which to base their identification. In discussing how their results might apply to actual eyewitnesses, they suggested that stereotypes about criminal appearance may be less influential when witnesses have encoded the culprit. This proposal, which we will test in Study 2, is in line with other research on stereotyping. For instance, stereotypes have less of an effect when decision makers have individuating information about a target, a result known as *stereotype dilution* (Locksley, Borgida, Brekke, & Hepburn, 1980; Locksley, Hepburn, & Ortiz, 1982).

In Study 2, participants attempted to identify a previously studied face from a lineup. The emotional appearance of the lineup suspect, whether "guilty" or "innocent," was the same as the previously studied face. For instance, if the participant studied the

"perpetrator" with an angry expression, then a lineup that had an angry suspect was shown to the participant at test. We held the emotional expression of the study face and the lineup suspect constant because, arguably, criminal face bias is most likely to occur when an eyewitness reports to the police that the culprit was angry, and then the police show the eyewitness a lineup in which the suspect is the only person who appears angry. After the learning phase, participants were given a target-present or a target-absent lineup. The fillers either matched or did not match the emotional appearance of the suspect. If criminal appearance drives which person in the lineup is identified, participants should tend to pick both guilty and innocent suspects more often if the suspect is the only person who appears angry. Put differently, participants will adopt a more lenient decision standard when there is only one face in the lineup that appears angry and more criminal-looking, thereby increasing the hit rate (i.e., the rate of guilty suspect identifications) in target-present lineups and increasing the false alarm rate (i.e., the rate of innocent suspect identifications) in target-absent lineups.

We also considered yet another way in which matching the fillers to the suspect's emotional appearance could affect memory retrieval in a lineup. Namely, matching the suspect and the fillers on emotional appearance could affect discrimination accuracy, or the ability to differentiate the guilty suspect from the fillers. Enhanced discrimination accuracy is demonstrated by an increased hit rate when the target is present and a decreased false alarm rate when the target is absent. Matching emotional appearance could affect discrimination accuracy in the following way: When the lineup members all appear angry, criminal appearance will not be a diagnostic cue for recognition because everyone in the lineup also has that feature. As a result, a witness will have to use additional features besides emotional appearance to recognize the culprit. The use of additional features in recognition should improve discrimination accuracy. Therefore, in this way, matching the fillers to the emotional appearance of the suspect could enhance discrimination accuracy compared to when the fillers do not match.

Study 2

Method

Design. A 2 (suspect emotional expression: low anger vs. high anger) \times 2 (filler type: matched vs. unmatched) \times 2 (target: present vs. absent) factorial design was employed. The outcome measures were decision confidence, which was measured on an 11-point Likert scale ranging from 0% *confident* to 100% *confident*, and identification outcome (select target, select filler, or reject the lineup).

Participants. A total of 1,983 adults were recruited from the general Internet population via Amazon Mechanical Turk, and were compensated 20 cents (USD) for their participation. Participants ranged in age from 18 to 75 years ($M = 30.22$, $SD = 9.63$ years; 37% female). The race/ethnic background of participants was 57% South Asian, 34% Caucasian, and 9% reported other backgrounds.

Materials and procedure. White males ($n = 40$) aged between 18 to 25 years were randomly selected from the Oklahoma Offender database. We used actual police mugshots as stimuli to

increase external validity. It is unlikely that the police are going to create lineups using stimuli that are as controlled as the images in the Radboud Faces Database. Our experience with police mugshots tells us that these images vary not only in the physical appearance of the persons shown, but also in the technical aspects of the photo itself (e.g., focal distance, position of the head, lighting). Hence, it is important to examine whether matching the fillers to the suspect on emotional appearance makes a difference when there are multiple features varying across lineup members, as is likely to be the case in actual police lineups. Additionally, the mugshots were rendered black and white to ensure that certain photos were not more memorable because of the fact that the colors were more vivid. We also edited the photos such that only the face and neck of the person were shown to eliminate cues to criminality that are unrelated to a person's face, such as jailhouse clothing or a height board, which were present in some of the photos.

Another reason why we used actual police mugshots was because they represent a range of facial features that might make a face in an actual police lineup appear criminal. The mugshots varied in a number of aspects that could affect perceptions of anger, from overt displays of anger, to the positioning of the head, which could affect the perceived distance between features. In terms of measuring perceptions of anger and criminality, we thought that it was important to capture this naturally occurring variation because, ultimately, this research may be generalized to the construction of police lineups. Thus, we carried out pilot work to select faces from the database that appeared the most, and the least, angry-looking. An independent group of participants ($n = 60$) rated each of the faces on either how angry or how criminal the face appeared using a 7-point Likert scale, ranging from *not at all [attribute]* to *very [attribute]*. Ratings were averaged across raters for each face. The seven faces that were rated the highest on anger were selected to serve as high anger faces (anger: $M = 4.67$, $SD = 0.68$; criminal: $M = 5.44$, $SD = 0.54$), and the seven faces that were rated the lowest on anger were selected to serve as low anger faces (anger: $M = 1.44$, $SD = 0.15$; criminal: $M = 2.62$, $SD = 0.65$).

We had participants study five faces, because in a real world crime, there is often more than one person present (e.g., the victim, bystanders). Consequently, the witness attends to more than one face, and this might negatively affect the accessibility of the memory trace for the culprit. One of the study faces was either a high anger face or a low anger face, and the other four study faces were in the middle of the face distribution with respect to anger and criminal appearance (anger: $M = 3.02$, $SD = 0.11$; criminal: $M = 3.91$, $SD = 0.52$). For each participant, the high or low anger face in the study set was randomly chosen from the set of high or low anger faces. The study faces were 50% larger in size compared to the lineup photos in order to reduce the likelihood that performance would be based on image-specific information (see *Kolers, Duchnick, & Sundstrom, 1985; Lee, Matsumiya, & Wilson, 2006; Longmore, 2007*).

We created a six-person simultaneous lineup (presented in a 2×3 array) for each participant. For a target present lineup, the high or low anger face that was shown during the study phase was in the lineup, and this face served as the target. For a target-present matched lineup, the fillers matched the emotional expression of the target. We randomly selected matched fillers from the faces that

remained in the set of high or low anger faces, depending on the target's emotional expression. For a target-present unmatched lineup, the fillers did not match the emotional expression of the target. For a high anger target in an unmatched lineup, five fillers were randomly selected from the remaining faces in the low anger face set. For a low anger target in an unmatched lineup, five fillers were randomly selected from the remaining faces in the high anger face set.

In the target-absent condition, the high or low anger face shown during the study phase was not displayed in the lineup. If the participant had studied a high anger face, we created a target-absent, matched lineup by selecting the six faces that remained in the set of high anger faces. One of the faces was randomly designated as the "innocent suspect." To create a target-absent, unmatched lineup for a participant who had studied a high anger face, we drew five faces from the low anger face set and one face from the high anger face set. The high anger face was the "innocent suspect." We followed these procedures to create target-absent lineups for the low anger suspect condition.

Procedure. Participants were told that they would be answering questions about images. Participants were first asked demographic questions and then they were shown the study faces, one at a time, for 3 seconds. After participants had engaged in a filler task (solving anagrams) for 5 minutes, they were given a surprise simultaneous lineup test. They were asked to examine the photos and report whether one of the previously studied faces was present. Participants were warned before seeing the lineup that a previously studied face may not be present and that not selecting a face may be the correct answer. Participants were given the option to not choose anyone or to say "don't know." Decision confidence was queried after the participants had made their lineup decision. Participants responded to only one lineup.

Results

Data analysis. We used Receiver Operating Characteristic (ROC) analysis to examine how well participants discriminated the target holding constant response criteria (*Macmillan & Creelman, 2005*; also see *Mickes, Flowe, & Wixted, 2012* for an application of ROCs to lineups). An ROC curve plots the cumulative hit and false alarm rates across the range of possible response criteria. Decision confidence established the response criterion the participant employed. An ROC curve was plotted for each experimental condition. We determined the number of participants who selected the suspect in the target present and target absent conditions at each level of confidence and then calculated the hit and false alarm rate at every confidence level. We plotted the ROC curves by cumulating the hit and false alarm rates across confidence levels, starting with the highest confidence level and ending with the lowest confidence level. The ROCs for the suspect when placed in a matched, in comparison to an unmatched, lineup for each suspect emotion condition are shown in *Figure 1*. The lineup response outcomes (collapsed across confidence level) by lineup condition and target condition, are provided in *Table 4*.

We undertook partial area under the curve (pAUC) analysis using pROC (*Robin et al., 2011*) to test whether discrimination accuracy varied across conditions. pAUC is appropriate because, typically, the false alarm rate for simultaneous lineups does not equal 1.00 (*Clark, Howell, & Davey, 2008; Steblay, Dysart, &*

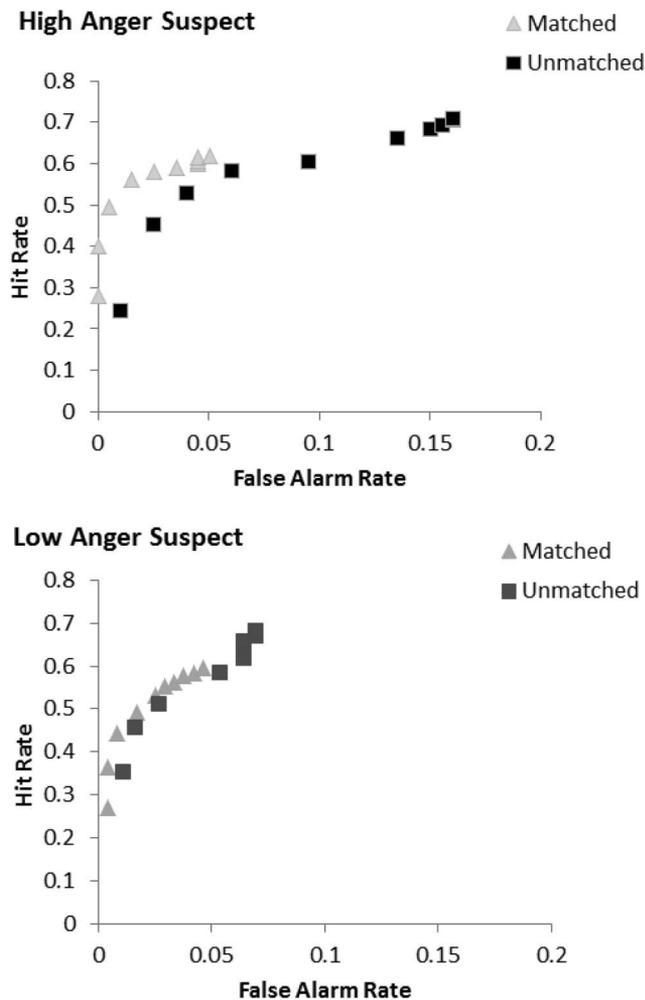


Figure 1. ROC curves for when the suspect is presented in an emotion-matched versus an emotion-unmatched lineup as a function of the suspect's emotional appearance.

Wells, 2011). We set pAUC to encompass the range of false alarm rates (i.e., the rate of “innocent suspect” identifications) that were obtained in the present study (see Table 5). The highest false alarm rate obtained across all of the conditions was 16%; therefore, specificity was set to 16% for the pAUC analysis. The difference between pAUCs is evaluated using the D statistic, defined, here, as $(\text{pAUC unmatched} - \text{pAUC matched})/s$, where s is the standard error of the difference between the two pAUCs estimated by the bootstrap method. The number of bootstraps was set to 10,000. If the pAUC analysis was not significant, we compared the hit and false alarm rates using chi square to test whether the response criteria varied across conditions.

High anger suspect identifications. Descriptively speaking, participants were more likely to accurately identify a high anger target that they had previously studied, and less likely to inaccurately select a high anger suspect that they had not studied, when the suspect was placed in a matched, compared to an unmatched, lineup (see the top panel of Figure 1). pAUC analysis confirmed that discrimination accuracy was greater in matched, compared to

unmatched, lineups (pAUC matched = .11 95% CI [.10–.12] and pAUC unmatched = .09, 95% CI [.07–.10]), $D = 2.33$, $p < .01$.

Low anger suspect identifications. Figure 1, bottom panel, provides the ROCs for the low anger suspects in matched, in comparison to unmatched, lineups. Discriminability was not significantly different across matched and unmatched lineups for low anger suspects, $D = 0.30$ (pAUC matched = .09, 95% CI [.09–.11] and pAUC unmatched = .10, 95% CI [.08–.11]), $p > .05$. Next, we examined whether participants employed a lower decision criterion in unmatched compared to matched lineups. The hit rate and the false alarm rate were higher in unmatched as opposed to matched lineups, which is consistent with participants having employed a more lenient criterion on average in the matched condition. While the innocent suspect was not selected significantly more often if presented in an unmatched ($p = .07$, 95% CI [.05, .12]) as opposed to matched lineup ($p = .04$, 95% CI [.03, .08]), $p > .05$, the guilty suspect was selected more often in an unmatched ($p = .69$, 95% CI [.62, .75]) compared to matched ($p = .59$, 95% CI [.54, .64]) lineup, $z = 2.24$, $p < .05$.

Remembering high versus low anger suspects. The results thus far indicate that matching the fillers to the suspect has a differential effect on remembering depending on the emotional expression of the suspect. In particular, discriminability was enhanced when the fillers matched the suspect's emotional expression, but only for a high anger suspect. However, perhaps differences in memorability, not emotional appearance, across the suspect conditions account for these results. To examine this possibility, we tested whether memory accuracy differed for high compared to low anger suspects. Our results indicated that there were no significant differences in discrimination accuracy across the suspect emotion conditions. When the fillers were matched to the emotional expression of the target, discrimination accuracy for high versus low anger targets did not significantly differ (pAUC high anger = .11, 95% CI [.10–.12] and pAUC low anger = .09, 95% CI [.08–.11]). When the fillers were unmatched, discrimination accuracy for high compared to low anger suspects did not significantly differ (pAUC high anger = .09, 95% CI [.07–.10] and pAUC low anger = .10, 95% CI [.08–.11]). Thus, the results suggest that matching the fillers to the emotional appearance of the suspect has a differential effect on memory retrieval depending on the suspect's emotional appearance.

General Discussion

The present study contributes to knowledge about the psychological mechanisms of eyewitness identification. According to the findings of Study 1, emotional appearance underpins identification decisions when a witness has no memory for a perpetrator and is compelled to identify someone. Under these conditions, the lineup member who appears relatively the most angry-looking is selected. Additionally, we found that angry faces tended to appear criminal-looking. This suggests that mock witnesses select angry faces because they are stereotypically criminal in appearance. Study 2 investigated whether angry suspects were differentially recognized depending on whether the fillers also appeared angry. We found that matching the fillers to the emotional appearance of the suspect improved people's ability to accurately identify the target. We will now consider these findings in further detail.

Table 5
Identification Outcomes by Suspect Emotion (High or Low Anger), Target Condition (Present or Absent) and Filler Emotional Appearance (Matched or Unmatched to the Suspect), Study 2

	Target-present				Target-absent				
	Target	Fillers	Reject	DK	Target	Fillers	Reject	DK	
HIGH ANGER SUSPECT									
Matched (<i>n</i> = 422)	0.62	0.22	0.12	0.04	Matched (<i>n</i> = 205)	0.05	0.35	0.53	0.07
Unmatched (<i>n</i> = 196)	0.68	0.11	0.17	0.04	Unmatched (<i>n</i> = 201)	0.16	0.33	0.41	0.10
LOW ANGER SUSPECT									
Matched (<i>n</i> = 346)	0.59	0.15	0.16	0.10	Matched (<i>n</i> = 246)	0.04	0.34	0.57	0.04
Unmatched (<i>n</i> = 179)	0.69	0.18	0.12	0.02	Unmatched (<i>n</i> = 188)	0.07	0.39	0.48	0.06

Note. DK = don't know.

Our results fit with the general face perception literature, which finds that perceptions of negative emotion underlie evaluations of trustworthiness (Oosterhof & Todorov, 2008; Said et al., 2009). We found in Study 1 and Study 2 that faces perceived as more angry-looking were also perceived as more criminal-looking. The results held even when the face was posed to display a happy or an emotionally neutral expression (Study 1). These findings are consistent with other research indicating that morphological characteristics of the face can make some faces appear emotional (Neth & Martinez, 2009; Said et al., 2009). It is interesting that these findings point to the conclusion that some suspects may naturally appear more angry-looking than others, even when they are trying to appear emotionally neutral or happy. Our research extends the face perception literature by showing that overt displays of emotion and morphological characteristics of faces can impact decision making in the context of lineups.

Our findings also extend the criminal face bias literature. Previously, it has been shown that mock witnesses will identify the most criminal-looking lineup member when they do not have a description of the culprit (Flowe & Humphries, 2011). We extended this finding by examining whether matching the fillers to the suspect's emotional appearance affected suspect identifications. This examination was carried out with a mock witness evaluation (Study 1) and a recognition memory test (Study 2). In Study 1, we found that selecting fillers who appeared angry rather than neutral in expression led to fewer mock witness identifications of angry suspects. In Study 2, we carried out a test of criminal face bias in the context of recognition memory. This was an important extension because, in other research, stereotypes do not seem to affect decision making when people have more relevant information on which to base their decision (e.g., Locksley et al., 1980, 1982). Our results indicated that participants did not simply select the person who appeared the most criminal-looking. Guilty and innocent suspects were not more likely to be identified when the suspect was the only member who appeared angry. What is more, recognition accuracy was increased by matching the fillers to the suspect on emotional appearance. Discrimination accuracy improved (i.e., the hit rate in the target-present condition was higher and the false alarm rate in the target-absent condition was lower) when the fillers matched the target's angry expression compared to when the expressions were different. Our interpretation of these findings is that when the emotional appearance of the fillers and the suspect match, participants have to engage in a more

extensive memory search. When all of the faces appear angry, participants examine additional features besides emotional appearance to determine whether a face is a sufficient match to the target in memory. When the target is the only face that has an angry appearance, the memory retrieval process may be less thorough, which leads to reduced accuracy.

It is interesting to note that the retrieval process for suspects who did not appear angry (i.e., low anger suspects) differed compared to angry suspects. In particular, the data for low anger suspects suggested that participants employed a more lenient decision standard when the fillers did not match the suspect compared to when they did match the suspect. The findings for low anger suspects are in line with other research on lineup member similarity, which finds that both guilty and innocent suspect identifications increase when the lineup members are low in similarity to one another (Clark, 2012; Fitzgerald, Price, Oriet, & Charman, 2013). An interesting question for further research is to examine why the retrieval process seems to differ for suspects who appear angry. In any case, we can conclude from the present studies that matching the fillers to an angry suspect's emotional appearance reduces the odds that the suspect is identified by chance alone (Study 1) and enhances memory retrieval (Study 2).

Applied Implications

One possible remedy for reducing criminal face bias might be to ensure that suspects make an emotionally neutral or a positive display when having their mugshot taken. This might not necessarily solve the problem, however. Our study, and others (Neth & Martinez, 2009; Said et al., 2009), shows that faces in a neutral or happy expression can be perceived as relatively angry-looking. Furthermore, if a happy or a neutral suspect was placed in a lineup with fillers who appeared relatively less angry in appearance, witnesses might still be inclined to select the suspect more often. In particular, we found that in a lineup in which all members were making an emotionally neutral display, suspects were selected more often the angrier and more criminal-looking they appeared. Thus, simply asking a person who is having their mugshot taken to make a neutral or positive emotional expression may not reduce criminal face bias for all suspects.

Another approach for possibly mitigating the influence of criminal face bias on lineup identifications is to select fillers that match the emotion perceived from the suspect's face as shown in the

lineup photograph. On the one hand, it might be argued that selecting the fillers based on their physical match to the suspect's emotional appearance may reduce the hit rate, because matching emotional expression will increase the physical similarity of the lineup members (see Luus & Wells, 1991). However, results are inconsistent across experiments (see Tunnicliff & Clark, 2000 for a review), with some showing higher hit rates for description-matched compared to suspect-matched lineups (Lindsay, Martin, & Webber, 1994; Luus & Wells, 1991) and others not finding any difference between the two filler selection strategies (Tunnicliff & Clark, 2000). The results for false alarms are also conflicting, with some finding no difference across strategies (Luus & Wells, 1991), and others finding that false alarm rates tend to be larger in description-matched compared to suspect-matched lineups (Lindsay et al., 1994; Tunnicliff & Clark, 2000). Given these inconsistencies, it seems premature to abandon the match to suspect strategy. What is more, our data indicate that matching the fillers to the emotional appearance of the suspect can enhance discrimination accuracy. Therefore, we think that further research examining the efficacy of various filler selection strategies to create fair lineups for angry suspects is warranted.

Limitations and Conclusion

There are at least three limitations of our results. First, whether criminal stereotypes are activated during a lineup test may depend on the type of crime that the eyewitness saw. As a starting point, we examined the effects of criminal stereotypes in an emotionally neutral context, wherein participants studied faces that were not embedded within a crime scenario. Although emotional appearance affected identifications even under these circumstances, the impact of criminal stereotypes might vary depending on the extent to which the crime is stereotypically associated with angry perpetrators. Second, the emotional appearance of the to-be-remembered face and the suspect face were the same. In our view, criminal face bias is most likely to manifest when an eyewitness reports to the police that the culprit was angry, and the police then show the eyewitnesses a lineup in which the suspect is the only person who appears angry. Therefore, we thought it was important to begin by investigating criminal face bias under these conditions. Further investigation will be required though to deepen our understanding of how the correspondence between the culprit's and the suspect's emotional appearance affects lineup recognition accuracy. Some basic face recognition research suggests that eyewitnesses who encoded an angry culprit might be less likely to identify the culprit if he or she does not also appear angry in the lineup. For instance, angry emotional expressions interfere with the processing of local facial features (Eastwood, Smilek, & Merikle, 2003) and reduce elaborative processing during the encoding of facial identity (D'Argembeau & Van der Linden, 2011). Therefore, a face encoded with an angry as opposed to a happy emotional expression is less likely to be recognized when it appears neutral at test (D'Argembeau & Van der Linden, 2011). Third, research is needed to examine the boundary conditions of the role that emotional appearance plays in lineup identifications. If an eyewitness has a particularly strong memory for the perpetrator, for example, recognition accuracy may be less affected by whether the fillers match the emotional expression of the suspect.

To summarize, our results indicate that emotional appearance plays an important role in the psychology of lineup identification. Faces that appear angry are perceived as more criminal-looking. Suspects who appear angry are more likely to be identified by mock witnesses if they are surrounded by emotionally neutral fillers rather than fillers who also appear angry. We further found that matching the fillers to an angry suspect's emotional appearance improved lineup recognition accuracy. Further research is needed to examine the best strategy for selecting fillers on emotional appearance.

References

- Becker, D. V., Kenrick, D. T., Neuberg, S. L., Blackwell, K. C., & Smith, D. M. (2007). The confounded nature of angry men and happy women. *Journal of Personality and Social Psychology, 92*, 179–190. doi:10.1037/0022-3514.92.2.179
- Bull, R. H. C., & Green, J. (1980). The relationship between physical appearance and criminality. *Medicine, Science, and the Law, 20*, 79–83.
- Bull, R., & McAlpine, S. (1998). Facial appearance and criminality. In A. Memon, A. Vrij, & R. Bull (Eds.), *Psychology and law: Truthfulness, accuracy and credibility* (pp. 59–76). London, UK: McGraw-Hill.
- Clark, S. E. (2012). Costs and benefits of eyewitness identification reform: Psychological science and public policy. *Perspectives on Psychological Science, 7*, 238–259. doi:10.1177/1745691612439584
- Clark, S. E., Howell, R., & Davey, S. D. (2008). Regularities of eyewitness identification. *Law and Human Behavior, 32*, 187–218. doi:10.1007/s10979-006-9082-4
- D'Argembeau, A., & Van der Linden, M. (2011). Influence of facial expression on memory for facial identity: Effects of visual features or emotional meaning? *Emotion, 11*, 199–202. doi:10.1037/a0022592
- Deutsch, F. M., LeBaron, D., & Fryer, M. M. (1987). What is in a smile? *Psychology of Women Quarterly, 11*, 341–352. doi:10.1111/j.1471-6402.1987.tb00908.x
- Doob, A. N., & Kirshenbaum, H. M. (1973). Bias in police lineups—Partial remembering. *Journal of Police Science and Administration, 1*, 287–293.
- Eastwood, J. D., Smilek, D., & Merikle, P. M. (2003). Negative facial expression captures attention and disrupts performance. *Perception & Psychophysics, 65*, 352–358. doi:10.3758/BF03194566
- Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological Science, 17*, 383–386. doi:10.1111/j.1467-9280.2006.01716.x
- Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior, 1*, 56–75. doi:10.1007/BF01115465
- Fitzgerald, R. J., Price, H. L., Oriet, C., & Charman, S. D. (2013). The effect of suspect-filler similarity on eyewitness identification decisions: A meta-analysis. *Psychology, Public Policy, and Law, 19*, 151–164. doi:10.1037/a0030618
- Flowe, H. D. (2012). Do characteristics of faces that convey trustworthiness and dominance underlie perceptions of criminality? *PloS ONE, 7*(6), e37253. doi:10.1371/journal.pone.0037253
- Flowe, H. D., & Humphries, J. E. (2011). An examination of criminal face bias in a random sample of police lineups. *Applied Cognitive Psychology, 25*, 265–273. doi:10.1002/acp.1673
- 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. doi:10.3758/BF03333904
- Grayson, C. E. (1998). When a 5 isn't a 5: Sex, prototypes, stereotypes and standards of comparison in emotion judgments of self and others. *Dissertation Abstracts International, 59*, 0912.

- Gronlund, S. D., Carlson, C. A., Dailey, S. B., & Goodsell, C. A. (2009). Robustness of the sequential lineup advantage. *Journal of Experimental Psychology: Applied*, *15*, 140–152. doi:10.1037/a0015082
- Kline, P. (2000). *Handbook of psychological testing* (2nd ed.). London, UK: Routledge.
- Kolers, P. A., Duchnick, R. L., & Sundstroem, G. (1985). Size in the visual processing of faces and words. *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 726–751. doi:10.1037/0096-1523.11.6.726
- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D., Hawk, S., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion*, *24*, 1377–1388. doi:10.1080/02699930903485076
- Lee, Y., Matsumiya, K., & Wilson, H. R. (2006). Size-invariant but viewpoint-dependent representation of faces. *Vision Research*, *46*, 1901–1910. doi:10.1016/j.visres.2005.12.008
- Lindsay, R. L., Martin, R., & Webber, L. (1994). Default values in eyewitness descriptions: A problem for the match-to-description lineup foil selection strategy. *Law and Human Behavior*, *18*, 527–541. doi:10.1007/BF01499172
- Locksley, A., Borgida, E., Brekke, N., & Hepburn, C. (1980). Sex stereotypes and social judgment. *Journal of Personality and Social Psychology*, *39*, 821–831. doi:10.1037/0022-3514.39.5.821
- Locksley, A., Hepburn, C., & Ortiz, V. (1982). Social stereotypes and judgments of individuals: An instance of the base-rate fallacy. *Journal of Experimental Social Psychology*, *18*, 23–42. doi:10.1016/0022-1031(82)90079-8
- Longmore, C. A. (2007). Learning faces from photographs. *Dissertation Abstracts International* (UMI No. 301710683).
- Luus, C. A. E., & Wells, G. L. (1991). Eyewitness identification and the selection of distracters for lineups. *Law and Human Behavior*, *15*, 43–57. doi:10.1007/BF01044829
- MacLin, M. K., Downs, C., MacLin, O. H., & Caspers, H. (2009). The effect of defendant facial expression on mock juror decision-making: The power of remorse. *North American Journal of Psychology*, *11*, 323–332.
- MacLin, M. K., & Herrera, V. (2006). The criminal stereotype. *North American Journal of Psychology*, *8*, 197–208.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.
- Macrae, N., & Shepherd, J. (1989). Stereotypes and social judgments. *British Journal of Social Psychology*, *28*, 319–325. doi:10.1111/j.2044-8309.1989.tb00875.x
- Malpass, R. S. (1981). Effective size and defendant bias in eyewitness identification lineups. *Law and Human Behavior*, *5*, 299–309. doi:10.1007/BF01044945
- Malpass, R. S., & Lindsay, R. C. L. (1999). Measuring lineup fairness. *Applied Cognitive Psychology*, *13*, S1–S7. doi:10.1002/(SICI)1099-0720(199911)13:1+<S1::AID-ACP678>3.0.CO;2-9
- McQuiston, D. E., & Malpass, R. S. (2002). Validity of the mock witness paradigm: Testing the assumptions. *Law and Human Behavior*, *26*, 439–453. doi:10.1023/A:1016383305868
- Mickes, L., Flowe, H. D., & Wixted, J. T. (2012). Receiver operating characteristic analysis of eyewitness memory: Comparing the diagnostic accuracy of simultaneous versus sequential lineups. *Journal of Experimental Psychology: Applied*, *18*, 361–376. doi:10.1037/a0030609
- Neth, D., & Martinez, A. M. (2009). Emotion perception in emotionless face images suggests a norm-based representation. *Journal of Vision*, *9*(1), article 5. doi:10.1167/9.1.5
- Olivola, C. Y., & Todorov, A. (2010). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *Journal of Experimental Social Psychology*, *46*, 315–324. doi:10.1016/j.jesp.2009.12.002
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 11087–11092. doi:10.1073/pnas.0805664105
- Porter, S., & ten Brinke, L. (2009). Dangerous decisions: A theoretical framework for understanding how judges assess credibility in the courtroom. *Legal and Criminological Psychology*, *14*, 119–134. doi:10.1348/135532508X281520
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*, 77–84. doi:10.1186/1471-2105-12-77
- Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion*, *9*, 260–264. doi:10.1037/a0014681
- 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. doi:10.1093/sf/51.4.427
- Stebly, N. K., Dysart, J. E., & Wells, G. L. (2011). Seventy-two tests of the sequential lineup superiority effect: A meta-analysis and policy discussion. *Psychology, Public Policy, and Law*, *17*, 99–139. doi:10.1037/a0021650
- Technical Working Group for Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement*. Washington, DC: National Institute of Justice. Retrieved from <https://www.ncjrs.gov/pdffiles1/nij/178240.pdf>
- Tredoux, C. G. (1998). Statistical inference on measures of lineup fairness. *Law and Human Behavior*, *22*, 217–237. doi:10.1023/A:1025746220886
- Tunnicliff, J. L., & Clark, S. E. (2000). Selecting foils for identification lineups: Matching suspects or descriptions? *Law and Human Behavior*, *24*, 231–258. doi:10.1023/A:1005463020252
- Valla, J. M., Ceci, S. J., & Williams, W. M. (2011). The accuracy of inferences about criminality based on facial appearance. *Journal of Social, Evolutionary, and Cultural Psychology*, *5*, 66–91. doi:10.1037/h0099274
- 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. doi:10.1002/(SICI)1099-0720(199911)13:1+<S27::AID-ACP635>3.3.CO;2-D
- 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. doi:10.1007/BF01039807
- Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior*, *22*, 603–647. doi:10.1023/A:1025750605807
- Yarmey, D. A. (1993). Stereotypes and recognition memory for faces and voiced of good guys and bad guys. *Applied Cognitive Psychology*, *7*, 419–431. doi:10.1002/acp.2350070505
- Zebrowitz, L. A., & McDonald, S. M. (1991). The impact of litigants' baby-facedness and attractiveness on adjudications in small claims courts. *Law and Human Behavior*, *15*, 603–623. doi:10.1007/BF01065855

Received May 29, 2013

Revision received June 3, 2014

Accepted June 10, 2014 ■