Journal of Experimental Psychology:
Human Perception and Performance
2014, Vol. 40, No. 4, 1526–1541

0096-1523/14/$12.00 http://dx.doi.org/10.1037/a0036732

The Facial Width-to-Height Ratio Shares Stronger Links With Judgments of Aggression Than With Judgments of Trustworthiness

Shawn N. Geniole and Danielle S. Molnar
Brock University

Justin M. Carré
Nipissing University

Cheryl M. McCormick
Brock University

Variation in the facial width-to-height ratio (face ratio) is associated with judgments of aggression and of trustworthiness made by observers when viewing men's faces. Although judgments of aggression and of trustworthiness are correlated, they represent distinct constructs. We thus investigated the hypothesis that judgments of aggression share stronger associations with the face ratio than judgments of trustworthiness, and that judgments of aggression mediate the link between the face ratio and trustworthiness. Across 4 separate studies, involving 129 observers rating subsets of 141 photographs (original photographs of individuals who provided consent for their use) of clean-shaven (65 faces), unshaved (22 faces), or digitized male faces (54 faces; digitized faces were creating using facial modeling software), this hypothesis was supported. The correlations between the face ratio and judgments of aggression were moderate to strong in all 4 studies (*r*s = .45 to .70). Reaction time was measured in Study 4: Participants judged aggression faster than trustworthiness; thus, temporal precedence also supports the hypothesis that aggression mediates the link between the face ratio and trustworthiness. Sensitivity to the face ratio may therefore be part of a perceptual mechanism specialized to assess aggressiveness rather than trustworthiness in others, likely because of the greater necessity for rapid judgments of aggressive potential than trustworthiness.

Keywords: social judgments, aggression, trustworthiness, dominance

The wisdom of the aphorism "choose your battles wisely" is obvious in conflicts that involve physical aggression and the potential for injury or death. Accurate assessments of the strength, toughness, or aggressiveness of another individual would facilitate decisions to defer or contend in aggressive interactions (Sell et al., 2009). There is evidence of accurate estimations of formidability guiding aggressive behavior across numerous species (reviewed in Blanchard, Griebel, Pobbe, & Blanchard, 2011; Taylor & Elwood, 2003). Further, static images are sufficient to enable accurate judgments; assessments of dominance in chimpanzees (Kramer, King, & Ward, 2011; Kramer & Ward, 2012), and of fighting ability in humans (Sell et al., 2009; Třebický, Havlíček, Roberts, Little, & Kleisner, 2013), based on facial and body photographs, are above chance accuracy. Although information about another's formidability and/or their behavioral intentions can be inferred from emotional expressions (Fridlund, 1994; McArthur & Baron, 1983), many studies have found evidence for accurate assessments of strength (Fink, Neave, & Seydel, 2007; Sell et al., 2009), potential for violence (Stillman, Maner, & Baumeister, 2010), dominance, power, and assertiveness (e.g., Berry, 1990), using photographs of faces posed in neutral expressions. Thus, the perceptual system seems to be tuned to cues in the face indicative of formidability and potential for aggressiveness.

One feature that may signal formidability and aggressive potential is the facial width-to-height ratio (face ratio), first described by Weston, Friday, and Liò (2007). Men with larger face ratios were more aggressive in laboratory tasks and in hockey games (i.e., had more penalty minutes per game; Carré & McCormick, 2008). Further, violent !Kung San men had wider faces than those who were nonviolent (Christiansen & Winkler, 1992), and professional mixed martial art fighters with a higher proportion of fight victories had wider faces than those with a lower proportion of fight victories (Třebický et al., 2013). Amygdala reactivity to threat, which is associated with aggression in clinical populations (reviewed in Coccaro, Sripada, Yanowitch, & Phan, 2011), shared a stronger link with self-reported aggression in men with larger rather than smaller face ratios (Carré, Murphy, & Hariri, 2013).

This article was published Online First May 12, 2014.

Shawn N. Geniole and Danielle S. Molnar, Department of Psychology, Brock University; Justin M. Carré, Department of Psychology, Nipissing University; Cheryl M. McCormick, Department of Psychology and the Centre for Neuroscience, Brock University.

We thank Kirstyn Ali, Amanda Keyes, and Klausa Manrique for assistance in data collection. We also thank Jonathan Simone for permission to use his photographs in the manuscript. The research was funded by a Social Sciences and Humanities Research Council (SSHRC) grant to CMM. CMM holds a Canada Research Chair in Behavioural Neuroscience and a Natural Sciences and Engineering Research Council (NSERC) Discovery Grant. SNG holds an SSHRC Canada Graduate Scholarship.

Correspondence concerning this article should be addressed to Cheryl M. McCormick, Department of Psychology and Centre for Neuroscience, Brock University, 500 Glenridge Ave., St. Catharines, Ontario, L2S 3A1 Canada. E-mail: cmccormick@brocku.ca

Although some have failed to replicate such effects using self-report measures of aggression (Özener, 2012) or reports of criminal history (Gómez-Valdés et al., 2013), self-reported behavior does not always predict actual behavior, and not all criminal acts involve aggression and violence. One study reported a marginal positive association ($p = .057$) between the face ratio and penalty minutes in professional hockey games (Deaner, Goetz, Shattuck, & Schnotala, 2012, p. 237), but this effect was later shown to be moderated by social status; the relationship was particularly pronounced among those lower, rather than higher, in socioeconomic status (Goetz et al., 2013).

In addition to behaving more aggressively, there is much evidence that men with larger face ratios are perceived to be more aggressive by observers than men with smaller face ratios. Specifically, observers rated men with larger face ratios as more aggressive than men with smaller face ratios (Boshyan, Zebrowitz, Franklin, McCormick, & Carré, 2013; Carré, McCormick, & Mondloch, 2009; Carré, Morrissey, Mondloch, & McCormick, 2010; Geniole, Keyes, Mondloch, Carré, & McCormick, 2012; Lefevre & Lewis, 2013; Short et al., 2012), even when other cues in the face related to masculinity, dominance, and strength (e.g., jaw width, forehead, lip size) were cropped from photographs (Carré et al., 2010). In one study that used photographs of faces of older men (M age = 52.5 years), however, the face ratio was associated with dominance rather than aggression (Alrajih & Ward, 2013). Although judgments of masculinity and of aggression are highly correlated, the face ratio predicted judgments of aggression independently of judgments of masculinity (Geniole et al., 2012). Judgments of aggressiveness of men's faces made by 8-year-old children in either Canada or China were correlated with the face ratio, indicating that the phenomenon appears early in development and may be universal (Short et al., 2012). Further, exaggerating the face ratio by tilting the head upward or downward (i.e., reducing the height but keeping the width constant) increased the extent to which faces appeared intimidating (Hehman, Leitner, & Gaertner, 2013). Therefore, the face ratio appears to be a cue of aggressiveness to which the human perceptual system is sensitive.

Several studies have found that the face ratio may be used for appraisals of trustworthiness (Efferson & Vogt, 2013; Kleisner, Priplatova, Frost, & Flegr, 2013). When face ratios were manipulated in photographs of men's faces, enhancements decreased observers' judgments of trust, whereas minimizations increased observers' judgments of trust (Stirrat & Perrett, 2010). Individuals with larger face ratios were more likely to exploit the trust of others for personal gain (Stirrat & Perrett, 2010), use explicit deception, and cheat in a lottery for a cash prize (Geniole, Keyes, Carré, & McCormick, 2014; Haselhuhn & Wong, 2012) than were those with smaller face ratios, indicating there is some accuracy to such judgments. One study, however, found no association between the face ratio and behavior in a trust game (Efferson & Vogt, 2013). Overall, the face ratio is found to predict the extent to which men's behavior is trustworthy or untrustworthy in addition to predicting aggressive behavior.

Irrespective of the accuracy of social judgments based on viewing faces, there is ample evidence of the consistency of such judgments across observers, and ample evidence that such judgments are relevant for social interactions (Todorov, Mende-Siedlecki, & Dotsch, 2013). There is much evidence that the face ratio is a cue in such judgments, with associations between the face ratio and judgments of aggression and trustworthiness in men among the highest associations reported for judgments based on facial cues, and with associations significant at the level of the individual rather than only at the level of the group (McCormick, 2013). Such consistency of judgments across individuals is notable in light of evidence that traits within the observer also influence judgments (e.g., M. L. Willis, Dodd, & Palermo, 2013). Determining the perceptual judgment to which the face ratio is most strongly and directly linked, trustworthiness or aggressiveness, would provide insight into two distinct, yet not necessarily independent, theoretical perspectives; that of the functional basis and evolution of the face ratio and that of the formation of social judgments.

With regard to the function and evolution of the face ratio, if variation in the face ratio is more strongly linked to judgments of aggression than to trustworthiness, then the face ratio may have functioned to reduce injury or the likelihood of death resulting from poor judgments of aggressive potential, and the consequential mismatching of opponents in aggressive interactions. Further, if the face ratio is more strongly linked to aggression than to trustworthiness, men with larger face ratios may have had more success than men with smaller ratios in achieving and maintaining dominance within social groups, by encouraging retreat and submission in aggressive encounters and by discouraging such challenges by other individuals. Consistent with this possibility, men with smaller face ratios were more likely to die from contact violence than were men with larger face ratios (Stirrat, Stulp, & Pollet, 2012). Conversely, if the face ratio is more strongly related to judgments of trustworthiness than to those of aggression, variation in this feature may have functioned to reduce the likelihood of exploitation and deception in social interactions (see Stirrat & Perrett, 2010). Men with larger, compared with smaller, face ratios would thus have benefited less in social groups, as they would be judged as less trustworthy and would have less opportunity for trade and collaboration. Therefore, knowing which judgment is more strongly associated with the face ratio is of theoretical importance, as it provides insight as to the function of, and maintenance of the variation in, this feature from an evolutionary perspective.

Clarifying the role of the face ratio in judgments of trustworthiness and aggression, and determining the order in which these judgments form, will also provide theoretical insight into the perceptual processes that shape such judgments. For example, although trustworthiness is considered a high-level cognitive judgment, trust judgments form rapidly (J. Willis & Todorov, 2006) and seem to rely on subcortical brain structures such as the amygdala (e.g., Adolphs, Tranel, & Damasio, 1998). One explanation for these counterintuitive findings is that when exposure to a face is limited, or the history of the individual is unknown (as in encounters with strangers), judgments of trustworthiness may "piggyback" on other social judgments (threat, aggressiveness, dangerousness) that are better suited for such circumstances.

There is some evidence that judgments of aggression may be better suited for rapid assessments during first encounters with strangers than are judgments of trustworthiness. For example, judgments of aggression appear to be more accessible, as they are provided at a frequency more than three times that of judgments of trustworthiness when describing faces (Oosterhof & Todorov, 2008). Further, although observers' confidence in trustworthiness

judgments decreased with shortened exposure time to face stimuli, their judgments of aggression did not (J. Willis & Todorov, 2006). Thus, it is possible that judgments of aggression are more relevant for circumstances in which exposure to a face is limited or in which the history of the individual is unknown. As such, judgments of aggression may be a mechanism through which judgments of trustworthiness form in such circumstances. Nevertheless, no studies, to date, have tested this "piggyback" framework.

Three predictions can be derived from the framework: (a) judgments of aggression should share stronger links with the face ratio than judgments of trustworthiness, (b) judgments of aggression should form faster than judgments of trustworthiness, and (c) judgments of aggression should mediate the link between the face ratio and judgments of trustworthiness. Although these three predictions have not yet been directly tested, one study provided indirect evidence that judgments of trustworthiness may form faster than judgments of aggression, which is contrary to our predictions (see the middle panel of Figure 2 in J. Willis & Todorov, 2006); the authors did not report a statistical test that compared the speed with which these two judgments were provided, however. Therefore, in the following studies, we tested the predictions derived from the "piggyback" framework: In Studies 1 through 3, we examined the strength of associations (both direct and indirect) between the face ratio and judgments of aggression and trustworthiness, and in Study 4, we examined the speed with which these two judgments were formed.

Study 1

Method

Participants. Participants (29 women, 5 men; *M* age = 20.06 years, *SD* = 3.67; 88% White, 12% other) were recruited through an online undergraduate research pool and received a $5 honorarium or a course credit for participation. All participants consented to the procedures of the study, which were approved by the Brock University Research Ethics Board.

Stimuli. Photographs were selected from a set of 74 men (*M* age = 20.16 years, *SD* = 2.78; 76% White, 24% other) who were photographed with a Nikon D50 digital camera while posing in a neutral facial expression and wearing a bouffant cap to conceal hairstyle. Of the 74 men, we only selected those who self-identified as White (to avoid any ethnicity-based stereotypes in judgments) and those who were facing the camera directly (e.g., some participants' heads were rotated such that face ratio measures were obscured). With these criteria, our sample of faces was reduced to 54 men (*M* age = 20.32 years, *SD* = 3.13).

Because the study was conducted during "Movember" (when men forsake shaving to promote prostate cancer awareness), many of the men photographed had facial hair, which may bias participants' judgments and/or obscure the relationships between the face ratio and these judgments (de Souza, Baumgasten, Baiao, & Otta, 2003; Kenny & Fletcher, 1973). Excluding individuals with facial hair reduced our sample to 25 men (*M* age = 19.52 years, *SD* = 1.69). The photographs of these men were standardized using the procedures described in Carré and colleagues (2009). Photos of the men with visible jewelry (e.g., earrings, necklace) were also modified using Adobe Photoshop to erase the visible jewelry and avoid observer bias in judgments.

To test our hypotheses on a larger set of stimuli than the set of 25 men without facial hair, we also created another set of faces by loading the larger sample of 54 male faces into Facegen, a 3-D facial modeling program (version 3.5; Singular Inversions, 2010). After loading the faces into Facegen, using the "PhotoFit" option, which involved placing landmarks on the pupils, left and right zygion, nostrils, lip corners, lower jaw extremities, and the chin of the male faces, the program estimated a 3-D model of the faces using methods similar to those described in Blanz and Vetter (1999). Textural details in each of the 54 faces were then removed using the "Detail Texture Modulation" analogue scale. This adjustment removed both the men's facial hair and blemishes (see Figure 1). After the adjustment, the Facegen faces were saved as 8-bit gray-scale bitmap images. Gray scale was used to minimize the influence of color tones in the face known to influence judgments of aggression (e.g., Stephen, Oldham, Perrett, & Barton, 2012). A paired samples *t* test indicated that the Facegen models of the faces had smaller face ratios than the original faces (original faces, $M = 1.90$, $SD = 0.17$; Facegen models, $M = 1.85$, $SD = 0.19$; $t[53] = -4.32$, $p < .001$), but the relative positions of each face within the distribution was maintained (the ratios of the Facegen faces were strongly correlated with the ratios of the original faces; $r = .89$, $p < .001$). Given that analyses examining the relationship between participants' judgments of the faces and the face ratio were performed within stimuli sets (real and Facegen stimuli sets), and not across stimuli sets, this difference in the relative size of the face ratios is not problematic. Other researchers have also provided evidence that social judgments share a similar factor structure if they are analyzed using real or Facegen faces (Oosterhof & Todorov, 2008). Thus, two sets of faces (25 clean-shaven faces, 54 Facegen faces) were used as stimuli for Study 1.

Measure of the face ratio. Face ratio measurements were calculated using the procedures outlined in Carré and McCormick (2008). Two naïve research assistants measured the face ratios. Interrater reliability of the face ratio measurements was high ($r = .94$, $n = 54$, $p < .001$).

Ratings of faces. The faces were presented for observers to rate using E-Prime software and a 17-in. Dell laptop monitor (approximately 15.2 × 12.9 visual degrees when viewed from

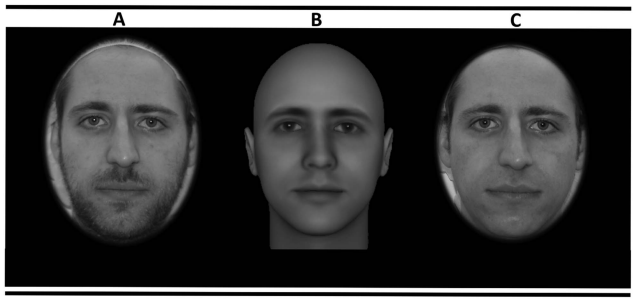

Figure 1. Panel A shows an example of a male face with facial hair. Panel B shows the same male after it was loaded into Facegen and the textural details were removed. Panel C shows the same male face after the man shaved, to allow for comparison with Panel B. The figure also shows how real faces and Facegen faces were presented to observers. Photographs are of Jonathan Simone, taken and altered by SNG, both of whom permit its inclusion in the manuscript.

75 cm). Before providing ratings for the 25 original clean-shaven and 54 Facegen versions of the faces, participants underwent a familiarization phase wherein they viewed (for 1,000 ms each) and rated six practice faces that were selected from a different set of stimuli. After the practice, participants rated the block of 25 real faces and the block of 54 Facegen faces, with the order of the blocks (real vs. Facegen) counterbalanced across participants. Within each block, a face was presented individually for 1,000 ms, after which a question and corresponding response scale appeared. For ratings of aggression, the question was, "How aggressive would this person be if provoked?" Responses were made using 7-point Likert scales (1 = *not at all*, 7 = *very much so*). Once the participant entered a response using a Dell Laptop standard keyboard (participants were given an unlimited time to make a response), the next photo appeared, and this process continued until all stimuli within a block were rated. Once all of the faces in a given block were rated on aggression, participants rated the same faces again for trustworthiness; the specific question was "How trustworthy does this person look?" Participants provided responses on the same 7-point Likert scale used for ratings of aggression.

The order of the presentation of faces was randomized across participants. After the entire block of original clean-shaven or Facegen version of faces was rated on both characteristics, a screen would appear asking participants to wait for the next set of instructions. At this point, participants completed a short demographic questionnaire and then started the remaining block of original or Facegen version of faces. Because of technical difficulties, one female participant could not complete the ratings of the Facegen faces. Thus, analyses on the original clean-shaven faces were conducted using mean ratings from 34 participants, whereas analyses on Facegen faces were conducted using mean ratings from 33 participants.

Statistical analyses. Pearson product–moment correlations were used to determine if ratings of aggression and of trustworthiness were associated with the face ratio for individual observers as well as for the group, as in our previous studies (McCormick, 2013). One-sample *t* tests were computed on the Fisher *z*-transformed correlations to test the hypothesis that the correlations of individual observers would be significantly different from zero. Linear regression was used to determine which judgments shared a unique association with the face ratio. Semipartial correlation coefficients (*sr*) from the linear regression are reported. Mediation models were conducted using aggregate data (based on averages; the mean rating provided by participants for each face), using bootstrapping with 5,000 samples of the aggregate data (Preacher & Hayes, 2008), and using multilevel modeling with HLM version 7.0 software (Raudenbush, Bryk, & Congdon, 2004). In multilevel modeling, every judgment is nested within each of the respective observers from which that judgment originated (e.g., Hehman, Leitner, Deegan, & Gaertner, 2013; Sell et al., 2009). This statistical approach utilizes all of the ratings made by each observer and is thus more sensitive to variability within observers. Sobel tests (Sobel, 1982) were conducted to assess the significance of indirect effects using an online calculator provided by Preacher and Leonardelli (2010). An alpha value of $p < .05$, two-tailed, was used to determine statistical significance.

Results

Relationships between the face ratio and ratings of aggression and trustworthiness: Analysis of correlations of individual observers.

Original, clean-shaven faces (25 faces). Judgments of aggression were significantly and positively correlated with the face ratio for 16 of the 34 observers ($M \pm 95\%$ confidence interval [CI] = .34 ± .10; one-sample *t* test, $t[33] = 6.94$, $p < .001$). Judgments of trustworthiness for the original clean-shaven faces were significantly and negatively correlated with the face ratio for 13 of the 34 observers ($M \pm 95\%$ CI = −.25 ± .11; one-sample *t* test, $t[33] = -4.57$, $p < .001$).

Facegen faces (54 faces). Judgments of aggression were significantly and positively correlated with the face ratio for 26 of the 33 observers ($M \pm 95\%$ CI = .37 ± .06; one-sample *t* test, $t[32] = 11.88$, $p < .001$). Judgments of trustworthiness were significantly and negatively correlated with the face ratio for 15 of the 33 ($M \pm 95\%$ CI = −.26 ± .08; one-sample *t* test, $t[32] = -7.11$, $p < .001$).

Is the relationship between the face ratio and ratings of trustworthiness mediated by ratings of aggression?

Original, clean-shaven faces (25 faces). The descriptive statistics for the face ratios and for the observers' judgments are provided in Table 1. The face ratio was positively correlated with the group mean ratings of aggression and negatively with the group mean ratings of trustworthiness for the original clean-shaven faces; aggression and trustworthy ratings were negatively correlated (see Figure 2a). To determine whether the relationship between the face ratio and judgments of trustworthiness were mediated by ratings of aggression, we used hierarchical linear regression with the face ratio entered on the first step, and ratings of aggression on the second step, as predictors of trustworthiness. If ratings of aggression mediate the relationship between the face ratio and trustworthiness, we would expect that the association between the face ratio and judgments of trustworthiness would no longer be significant once ratings of aggression were added as a predictor (Baron & Kenny, 1986). The results were consistent with our hypothesis (see Figure 2a). Furthermore, adding ratings of aggression accounted for significantly more variability in the ratings of trustworthiness ($\Delta R^2 = .30$, $p < .01$) than did the face ratio alone. The variance inflation factor (VIF) for this analysis was 1.61, indicating that multicollinearity likely did not obscure the results (many statisticians and researchers have suggested that VIFs greater than 10 are indicative of multicollinearity problems; Hair, Anderson, Tatham, & Black, 1995; Mason, Gunst, & Hess, 1989; Neter, Wasserman, & Kutner, 1989). Therefore, results indicate that individuals with larger face ratios are rated as less trustworthy because they are perceived as more aggressive than are individuals with smaller face ratios.

To ensure that ratings of aggression were more strongly or uniquely associated with the face ratio than were ratings of trustworthiness, we entered both as simultaneous predictors of the face ratio. Results indicated that aggression ($t = 2.37$, $\beta = .57$, $sr = .40$, $p = .03$), but not trustworthiness ($t = -0.24$, $\beta = -.06$, $sr = -.04$, $p = .81$), was a significant predictor of variation in the face ratio (VIF = 2.10).

Analyses using a bootstrapping technique with 5,000 samples (Preacher & Hayes, 2008) provided 95% CI (bias corrected) for the

Table 1
Descriptive Statistics for the Face Ratio and for Judgments Provided by Observers

Measure	Mean (*SD*)	Cronbach's α
Study 1: Faces without facial hair (n = 25)		
Face ratio	1.87 (0.16)	
Ratings of aggression	3.92 (0.78)	.92
Ratings of trustworthiness	3.85 (0.81)	.92
Study 1: Facegen faces (n = 54)		
Face ratio	1.85 (0.19)	
Ratings of aggression	4.00 (0.83)	.94
Ratings of trustworthiness	3.88 (0.71)	.90
Study 2: Faces with facial hair (n = 22)		
Face ratio	1.88 (0.15)	
Ratings of aggression	3.85 (0.79)	.88
Ratings of trustworthiness	4.15 (0.57)	.75
Ratings of hairiness	3.05 (1.54)	.99
Study 4: Faces without facial hair (n = 65)		
Face ratio	1.91 (0.16)	
Ratings of aggression	3.56 (0.87)	.93
Ratings of trustworthiness	3.51 (0.82)	.93

indirect relationship between the face ratio and judgments of trustworthiness (i.e., controlling statistically for judgments of aggression). Confidence intervals that do not overlap with a value of zero are indicative of mediation. Results from theses analyses were consistent with our original findings; specifically, the face ratio shared an indirect association with judgments of trustworthiness (95% CI [−3.93, −0.79]).

When the data were analyzed using multilevel modeling, the average correlations between the face ratio and judgments of trustworthiness ($\gamma = -2.38$, $SE = 0.47$, $t[815] = -5.05$, $p < .001$), and between the face ratio and judgments of aggression ($\gamma = 3.01$, $SE = 0.46$, $t[815] = 6.55$, $p < .001$), were significant. The average association between judgments of aggression and judgments of trustworthiness was also significant ($\gamma = -0.37$, $SE = 0.06$, $t[815] = -6.52$, $p < .001$). The average relationship between the face ratio and trustworthiness was attenuated, but nevertheless was still significant ($\gamma = -1.43$ $SE = 0.37$, $t[814] = -3.89$, $p < .001$), when judgments of aggression ($\gamma = -.32$, $SE = 0.05$, $t[814] = -5.91$, $p < .001$) was added as a simultaneous predictor, indicating partial mediation (see Table 2 for Sobel test).

Facegen faces (54 faces). The descriptive statistics for the face ratios and for the observers' judgments are provided in Table 1. The face ratio was positively correlated with the mean group ratings of aggression and negatively with the mean group ratings of trustworthiness; aggression and trustworthy ratings were negatively correlated (see Figure 2b). The relationship between the face ratio and ratings of trustworthiness was no longer significant when ratings of aggression were added to the model (see Figure 2b). Further, ratings of aggression accounted for significantly more variability in the ratings of trustworthiness ($\Delta R^2 = .49$, $p < .001$) than did the face ratio alone (VIF = 1.58). Therefore, results indicated that individuals with larger face ratios are rated as less trustworthy because they are perceived as more aggressive than are individuals with smaller face ratios. When both judgments were entered as simultaneous predictors of the face ratio to determine which shared a stronger and more unique association with the face ratio, aggression ($t = 2.77$, $\beta = .63$, $sr = .31$, $p < .01$), but not trustworthiness ($t = 0.14$, $\beta = .03$, $sr = .02$, $p = .89$), was a significant predictor (VIF = 4.23).

Results from the bootstrapped mediation also indicated that the face ratio shared an indirect association with judgments of trustworthiness (95% CI [−2.84, −1.36]). Using multilevel modeling, the average correlations between the face ratio and judgments of trustworthiness ($\gamma = -1.93$, $SE = .30$, $t[1,748] = -6.54$, $p < .001$), and between the face ratio and judgments of aggression ($\gamma = 2.61$, $SE = 0.23$, $t[1,748] = 11.31$, $p < .001$), were significant. The average association between judgments of aggression and of trustworthiness was also significant ($\gamma = -0.35$, $SE = 0.04$, $t[1,748] = -8.38$, $p < .001$). The average relationship between the face ratio and trustworthiness was attenuated, but still significant ($\gamma = -1.15$, $SE = 0.23$, $t[1,747] = -4.93$, $p < .001$), when judgments of aggression ($\gamma = -0.30$, $SE = 0.04$, $t[1,747] = -8.03$, $p < .001$) was added as a simultaneous predictor, indicating partial mediation (see Table 2 for Sobel test).

Study 2

To replicate and extend the findings of Study 1, we recruited a new set of observers and created a new set of stimuli using the photographs of men with facial hair from Study 1. If sensitivity to the face ratio is part of an evolved mechanism for assessing formidability in others, observers' judgments of aggression should be associated with the face ratio even when men have facial hair. We tested this hypothesis and also determined if ratings of aggression mediated the relationship between the face ratio and trustworthiness in this new subset of faces.

Method

Participants. Participants (12 women, 12 men; M age = 19.58, $SD = 1.47$, age range = 18 to 23 years; 92% White, 8% other) were recruited through an online undergraduate research pool and received a $5 honorarium or a course credit for participation. All participants consented to the procedures of the study, which were approved by the Brock University Research Ethics Board.

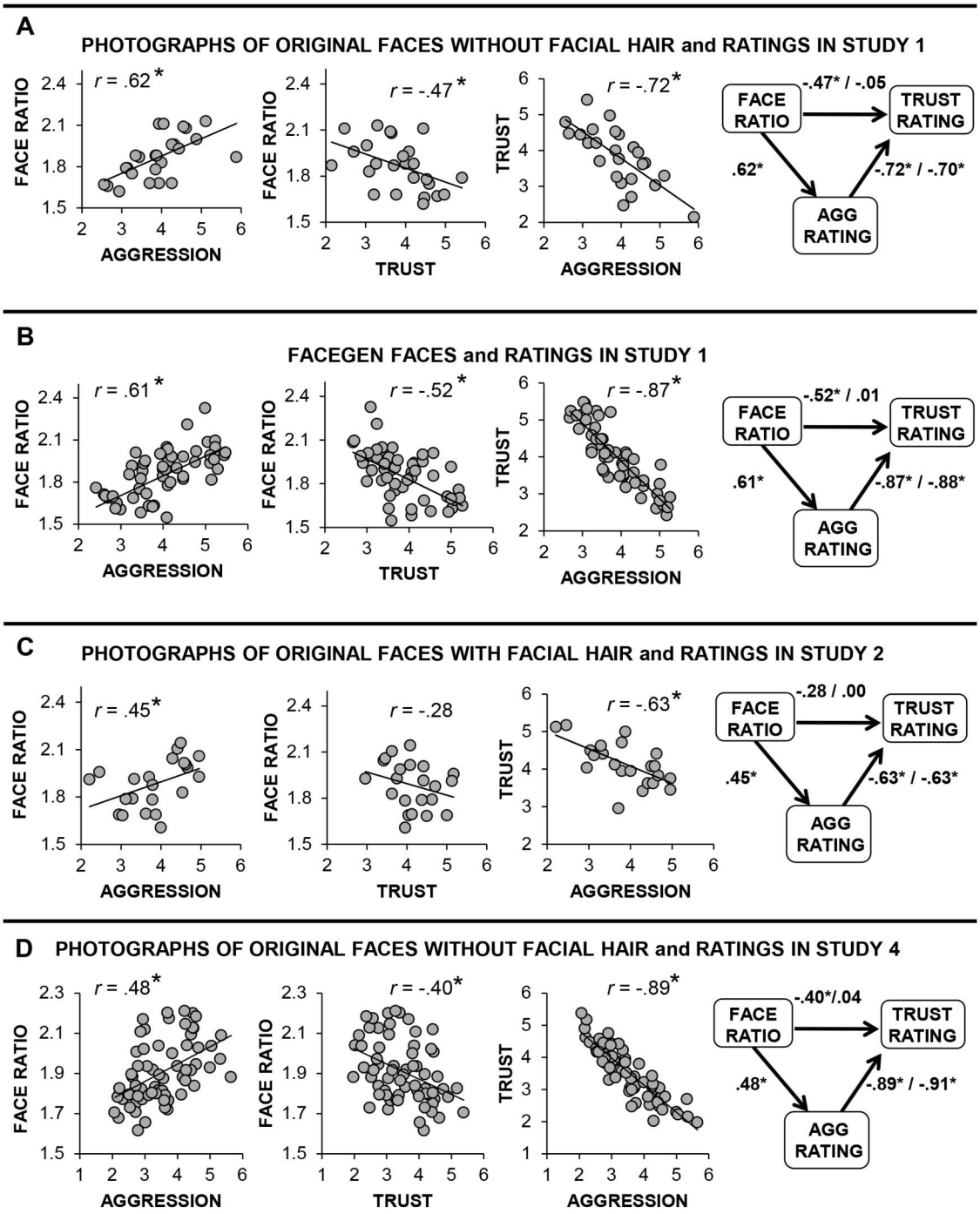

Figure 2. Scatterplots of the face ratios and observers' mean judgments of aggression and trustworthiness are displayed in the three left-most figures of panels A, B, C, and D. The right-most figure in each panel shows the results of mediation analyses used to determine whether the relationships between the face ratio and judgments of trustworthiness were mediated by judgments of aggression (AGG). The numbers shown are standardized regression coefficients (β weights). The first β weights between the face ratio and judgments of trustworthiness, and between judgments of aggression and judgments of trustworthiness, represent the strength of the bivariate relationships between these variables. The second β weights represent the strength of these relationships when the face ratio and judgments of aggression were entered as simultaneous predictors of judgments of trustworthiness.

Table 2
Sobel Tests of Mediation Across Studies

Study	Sobel test statistic	*p* (two-tailed)
Analyses using mean ratings across observers		
Study 1 (25 real faces)	2.64	.008
Study 1 (54 Facegen faces)	4.84	<.001
Study 2 (22 bearded faces)	1.85	.06
Study 3 (24 real faces; reanalysis of Carré et al., 2009)	3.18	.001
Study 4 (65 real faces)	4.18	<.001
Analyses using multilevel modelling		
Study 1 (25 real faces)	4.39	<.001
Study 1 (54 Facegen faces)	6.55	<.001
Study 2 (22 bearded faces)	2.84	.004
Study 3 (24 real faces; reanalysis of Carré et al., 2009)	8.13	<.001
Study 4 (65 real faces)	6.14	<.001

Note. Significant values indicate that aggression reduced the association between the face ratio and trustworthiness.

Stimuli. Only the faces of men with a significant amount of facial hair (i.e., more than stubble; $n = 22$) were selected from the set of 54 faces described in Study 1. These photos were also standardized using the procedures described in Carré and colleagues (2009).

Procedure. Participants completed demographic questionnaires and rated the 22 bearded male faces, first on trustworthiness and second on aggression. After rating the faces on aggression, participants provided ratings of hairiness based on the question, "How much facial hair does this person have?" Responses were provided using a 7-point Likert scale, in which 1 = *some* (*a little*) and 7 = *much* (*a lot*). Aside from changing the order of ratings, adding a rating of hairiness, and having only one block of faces, all procedures were the same as Study 1.

Results

Relationships between the face ratio and ratings of aggression and trustworthiness in men with facial hair: Analysis of correlations of individual observers. Because of the smaller sample of faces used in this study ($n = 22$), a stronger correlation was required to reach statistical significance than was required with the larger samples of faces. Nevertheless, four of the 24 observers provided judgments of aggression that were significantly correlated with the face ratio ($M \pm 95\%$ CI $= .23 \pm .11$; one-sample t test, $t[23] = 4.36$, $p < .001$), whereas only one of the 24 observers provided judgments of trustworthiness that were significantly associated with the face ratio ($M \pm 95\%$ CI $= -.12 \pm .09$; one-sample t test, $t[23] = -2.86$, $p < .01$).

Is the relationship between the face ratio and ratings of trustworthiness mediated by ratings of aggression in men with facial hair? The descriptive statistics for the face ratios and for the observers' judgments are provided in Table 1. When ratings were averaged across observers, judgments of aggression were negatively correlated with judgments of trustworthiness, but the face ratio shared a significant association only with judgments of aggression (see Figure 2c). Men ($n = 22$) who were rated as having more facial hair had larger face ratios ($r = .44$, $p = .04$) and were judged as more aggressive ($r = .46$, $p = .03$) and less trustworthy ($r = -.57$, $p < .01$) than were men rated as having less facial hair. Although the face ratio shared some negative association with judgments of trustworthiness (albeit nonsignificant, $p = .20$), this relationship was completely attenuated when judgments of aggression were added to the model (see Figure 2c). Further, judgments of aggression accounted for significantly more variability in the judgments of trustworthiness ($\Delta R^2 = .32$, $p < .01$) than did the face ratio alone (VIF = 1.26). When judgments of trustworthiness and of aggression were entered as simultaneous predictors of the face ratio to determine which variable shared a stronger unique association with the face ratio, judgments of aggression approached significance ($t = 1.73$, $\beta = .46$, $sr = .35$, $p = .10$), but those of trustworthiness did not ($t = 0.02$, $\beta = .01$, $sr = .00$, $p = .98$; VIF = 1.66).

Results from the bootstrapped mediation also indicated that the face ratio shared an indirect association with judgments of trustworthiness (95% CI [−2.20, −0.36]). Using multilevel modeling, the average correlations between the face ratio and judgments of trustworthiness ($\gamma = -1.06$, $SE = 0.35$, $t[503] = -3.05$, $p = .002$), and between the face ratio and judgments of aggression ($\gamma = 2.33$, $SE = 0.53$, $t[503] = 4.37$, $p < .001$), were significant. The average association between judgments of aggression and trustworthiness was also significant ($\gamma = -0.24$, $SE = 0.06$, $t[503] = -3.99$, $p < .001$). The average relationship between the face ratio and trustworthiness was attenuated ($\gamma = -0.53$, $SE = 0.37$, $t[502] = -1.43$, $p = .15$) when judgments of aggression ($\gamma = -0.23$, $SE = 0.06$, $t[502] = -3.74$, $p < .001$) was added as a simultaneous predictor, indicating mediation (see Table 2 for Sobel test).

Study 3

As an additional test of our hypotheses, we also analyzed data from a previously published study (Experiment 1 of Carré et al., 2009) in which 31 participants (16 women, 15 men) rated aggression, trustworthiness, and several other characteristics for facial photographs of 24 different men than those used in the present studies.

Method

Participants. Thirty-one undergraduates (16 women, 15 men) were recruited from Brock University and received course credit for participation (see Carré et al., 2009). Procedures were approved by Brock University's Research Ethics Board.

Stimuli. Study 3 involved the reanalysis of a previously published study (Study 1 of Carré et al., 2009), in which 24 male, clean-shaven, Caucasian faces were rated. In brief, the faces were standardized with a 400-pixel hairline to chin distance, elliptically cropped so that only the face was visible, and converted to 8-bit gray scale.

Procedure. Faces appeared for 2,000 ms, after which a question appeared. The question for aggression was, "How aggressive would this person be if provoked?" and the question for trustworthiness was, "How trustworthy does this person look?" Responses were provided using 7-point Likert scales similar to those used for aggression and trustworthiness in Study 1 and Study 2. For more details regarding the procedures, see Carré and colleagues (2009).

Results

There were significant bivariate correlations between ratings of aggression and of trustworthiness ($r = -.90$, $p < .001$), and between the face ratio and both of these ratings (face ratio and aggression, $r = .59$, $p < .01$; face ratio and trustworthiness, $r = -.45$, $p = .03$). Nevertheless, the face ratio was no longer a significant predictor of judgments of trustworthiness ($t = 1.16$, $\beta = .13$, $sr = .11$, $p = .26$) when judgments of aggression were added with the face ratio as a predictor into a hierarchical linear regression model. Furthermore, adding ratings of aggression accounted for significantly more variability in the ratings of trustworthiness ($\Delta R^2 = .63$, $p < .001$) than did the face ratio alone (VIF = 1.53). When both judgments were added as simultaneous predictors of the face ratio, ratings of aggression ($t = 2.53$, $\beta = 1.00$, $sr = .43$, $p = .02$) were a significant predictor of the face ratio, but those of trustworthiness were not ($t = 1.16$, $\beta = .46$, $sr = .20$, $p = .26$; VIF = 5.37).

Results from the bootstrapped mediation also indicated that the face ratio shared an indirect association with judgments of trustworthiness (95% CI [−5.35, −1.46]). Using multilevel modeling, the average correlations between the face ratio and judgments of trustworthiness ($\gamma = -2.41$, $SE = 0.40$, $t[703] = -5.98$, $p < .001$), and between the face ratio and judgments of aggression ($\gamma = 4.03$, $SE = 0.25$, $t[703] = 15.87$, $p < .001$), were significant. The average association between judgments of aggression and trustworthiness was also significant ($\gamma = -0.40$, $SE = 0.03$, $t[703] = -11.94$, $p < .001$). The average relationship between the face ratio and trustworthiness was attenuated, but still significant ($\gamma = -0.92$, $SE = 0.44$, $t[702] = -2.10$, $p = .04$), when judgments of aggression ($\gamma = -0.37$, $SE = 0.04$, $t[702] = -9.46$, $p < .001$) was added as a simultaneous predictor, indicating partial mediation (see Table 2 for Sobel test).

Study 4

The goal of Study 4 was to examine the speed with which participants rate faces on aggression compared with trustworthiness. If judgments of aggression are more relevant to survival than are judgments of trustworthiness, they should form more quickly than judgments of trustworthiness. Such a difference in the speed with which these judgments are provided would establish temporal precedence and bolster the hypothesis that aggression mediates the link between the face ratio and judgments of trustworthiness. This prediction was tested with a new set of observers and a larger set of stimuli.

Study 4 also had a few methodological advantages compared with Studies 1 through 3. Specifically, ratings were counterbalanced in Study 4 such that half of participants rated aggression first (and trustworthiness second), whereas the other half rated trustworthiness first (and aggression second). Although the consistency in our results from Studies 1 through 3, irrespective of the order of ratings (aggression rated first in Studies 1 and 3; trustworthiness rated first in Study 2), suggest that order is not an important factor, this conclusion may be limited because a different set of faces was used in each study. Therefore, Study 4 was better designed to account for potential order effects through the use of counterbalancing.

Study 4 also allowed for a tighter comparison between judgments of aggression and trustworthiness than did Studies 1 through 3 because the question used to prompt participants' judgments of aggression was rephrased ("How aggressive would this person be if provoked?" was changed to "How aggressive does this person look?") to parallel the question used for judgments of trustworthiness ("How trustworthy does this person look?").

Method

Participants. Participants (32 women, 8 men; *M* age = 19.38, *SD* = 1.86, age range = 17 to 26 years; 90% White, 10% other) were recruited through an online undergraduate research pool and received a $5 honorarium or a course credit for participation. All participants consented to the procedures of the study, which were approved by the Brock University Research Ethics Board.

Stimuli. To maximize the number of faces used in the analysis, the clean-shaven faces from Study 1 ($n = 25$) and from Carré and colleagues (2009; $n = 24$) were combined to form a larger set of faces. In addition, 16 photographs of cleanly shaven Caucasian male faces (collected during an ongoing, unrelated study; *M* age = 19.27 years, *SD* = 1.49) were added, for a total of 65 faces.

Measure of the face ratio. Face ratio measurements were calculated using the procedures outlined in Carré and McCormick (2008). Two naïve research assistants measured the face ratios. Interrater reliability of the face ratio measurements was high ($r = .88$, $n = 65$, $p < .001$).

Procedure. Participants were asked to judge aggression and trustworthiness as fast as possible using their "gut instincts," based on J. Willis and Todorov (2006). All participants made judgments of both characteristics, with the order of the judgments counterbalanced across participants. Before providing these judgments, however, participants completed a task to gauge their general response speed. This task was used to ensure that any differences in response time between those who rated aggression versus trust first were specific to these judgments, and not group differences in overall response speed. To assess general response speed, participants judged the size of circles as quickly as possible, using keys 1 through 7 on a laptop keyboard (1 = *smallest*, 7 = *largest*). There were seven different circle sizes and each size was presented

four times (with the order of presentation randomized across participants), for a total of 28 trials. The circles were white and were presented in the center of the computer screen, which had a black background. A white fixation cross appeared in the center of the screen for 500 ms before the presentation of each circle. After the circle appeared, it remained on the screen until a response was entered. After a response was entered, the fixation cross reappeared and the process repeated until each circle was rated.

After participants rated the size of the circles, they provided judgments of aggression and trustworthiness (order counterbalanced). The specific questions were "How aggressive does this person look?" (1 = *not at all aggressive*, 7 = *very aggressive*) and "How trustworthy does this person look?" (1 = *not at all trustworthy*, 7 = *very trustworthy*). Before providing judgments, participants were again reminded to use their "gut instincts" and to provide the responses as quickly as possible. During the task, a white fixation cross appeared for 500 ms, before the presentation of each face. Once the face was presented, it remained on the screen until participants provided a response. After providing a response, the fixation cross reappeared for 500 ms, and then another face was presented. This process repeated until each of the 65 faces was rated on either aggression or trustworthiness. After finishing the block of aggression, or of trustworthiness, participants read instructions about the next judgment they would provide and the same process was repeated.

Results

General response speed. A paired-samples *t* test on the average response times to each of the seven circle sizes indicated that the two groups did not differ in general response speed (aggression first: *M* ms = 1,048.71, *SD* = 66.05; trustworthiness first: *M* ms = 1,070.53, *SD* = 100.76; paired-sample *t* test, $t[6] = 0.91$, $p = .40$; see Figure 3a).

Do participants judge aggression faster than trustworthiness? To reduce the influence of response times that reflect lapses in attention, and to avoid the complete removal of genuine response times that may reflect increased difficulty in the formation of trust or aggression judgments, lengthy response times were Winsorized such that times longer than 4,000 ms were changed to 4,000 ms (greater than 2.5 *SD*s of the mean, 1.81% of aggression and 2.77% of trustworthiness response times).[1] The process and benefits of Winsorizing are discussed in many articles (e.g., Erceg-Hurn & Mirosevich, 2008; Ruppert, 1988; Wilcox, 2005) and Winsorizing has been used for reaction time (RT) outliers in many recent studies (e.g., Chambers, Swan, & Heesacker, 2014; Lai et al., 2012; Mueller, Makeig, Stemmler, Hennig, & Wacker, 2011; Townsend, Eliezer, Major, & Mendes, 2014; Wilkowski & Meier, 2010). A mixed factorial ANOVA with one within-subjects factor (aggression vs. trustworthiness judgments) and one between-subjects factor (aggression rated first or second) revealed two main effects: Participants rated aggression faster than trustworthiness ($F[1, 128] = 29.70$, $p < .001$, Cohen's $d = 0.96$), irrespective of whether they rated aggression or trustworthiness first, and participants provided both judgments faster if they rated aggression first rather than second ($F[1, 128] = 9.03$, $p < .01$, Cohen's $d = 0.53$; see Figure 3b). The interaction term was not significant ($F[1, 128] = 0.06$, $p = .81$). To allow us to determine which individuals were significantly faster to make one rating compared with the other, mean response times for aggression and for trustworthiness judgments were compared within each of the 40 observers using paired-samples *t* tests (see Figure 3c). Fourteen of the participants rated aggression significantly faster than trustworthiness, whereas three participants rated trustworthiness significantly faster than aggression ($ps < .05$). A chi-square test confirmed that this difference in the proportion of participants that rated aggression faster than trustworthiness was statistically significant ($\chi^2_1 = 7.12$, $p < .01$).

As another test of the hypothesis that judgments of aggression form faster than judgments of trustworthiness, a difference score was created for each participant by subtracting the average time to rate each face on trustworthiness from the average time to rate each face on aggression. A one-sample *t* test comparing these values with zero indicated that judgments of aggression were provided significantly faster than were judgments of trustworthiness ($M = 105.02$, $SD = 246.83$; $t[39] = 2.69$, $p = .01$, Cohen's $d = 0.86$).

Relationships between the face ratio and ratings of aggression and trustworthiness: Analysis of correlations of individual observers. Judgments of aggression were significantly and positively correlated with the face ratio for 24 of the 40 observers ($M \pm 95\%$ CI = $.26 \pm .06$; one-sample *t* test, $t[39] = 9.58$, $p < .001$). Judgments of trustworthiness for this same group were significantly and negatively correlated with the face ratio for 17 of the 40 observers ($M \pm 95\%$ CI = $-.21 \pm .04$; one-sample *t* test, $t[39] = -10.72$, $p < .001$).

Is the relationship between the face ratio and ratings of trustworthiness mediated by ratings of aggression? The descriptive statistics for the face ratios and for the observers' judgments are provided in Table 1. The face ratio was positively correlated with the mean ratings of aggression and negatively with the mean ratings of trustworthiness; aggression and trustworthy ratings were negatively correlated (see Figure 2d). The relationship between the face ratio and ratings of trustworthiness, however, was no longer significant when ratings of aggression were added to the model (see Figure 2d). Further, ratings of aggression accounted for more variability in the ratings of trustworthiness ($\Delta R^2 = .63$, $p < .001$) than did the face ratio alone (VIF = 1.31). When both judgments were entered as simultaneous predictors of the face ratio, judgments of aggression were significant ($t = 2.49$, $\beta = .60$, $sr = .28$, $p = .02$), but judgments of trust were not ($t = 0.55$, $\beta = .13$, $sr = .06$, $p = .58$; VIF = 4.77).

[1] The same results were obtained when lengthy response times (>4,000 ms) were trimmed (i.e., removed from the analysis). Specifically, when data were trimmed, a mixed factorial ANOVA with one within-subjects factor (aggression vs. trustworthiness judgments) and one between-subjects factor (aggression rated first or second) revealed two main effects: Participants rated aggression faster than trustworthiness ($F[1, 128] = 29.70$, $p < .001$, Cohen's $d = 0.96$), irrespective of whether they rated aggression or trustworthiness first, and participants rated both judgments faster if they rated aggression first rather than second ($F[1, 128] = 4.37$, $p = .04$, Cohen's $d = 0.37$). The interaction term was not significant ($F[1, 128] = 2.32$, $p = .13$). The 40 paired-samples *t* tests comparing the mean of aggression and the mean of trustworthiness response times within each of the 40 observers revealed that 15 of the participants rated aggression significantly faster than trustworthiness, whereas four participants rated trustworthiness faster than aggression ($ps < 0.05$).

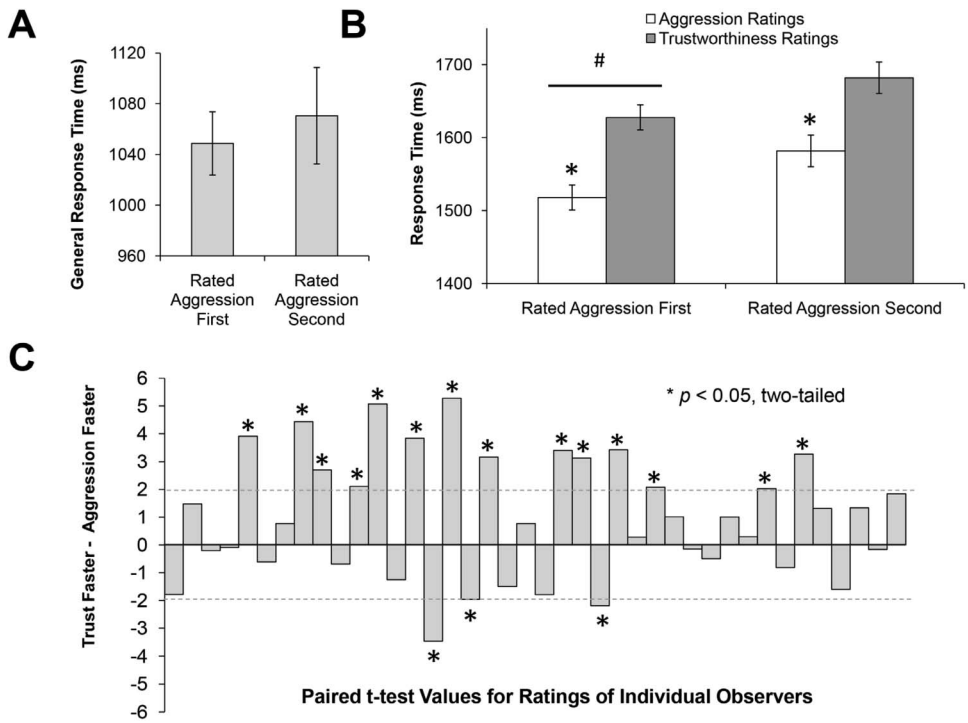

Figure 3. Panel A shows the participants' mean response speed in Study 4 when asked to rate the size of circles as quickly as possible. The bar on the left represents the mean and standard error of measurement (*SEM*) of participants who were asked to judge aggression first (and trustworthiness second), whereas the bar on the right represents the mean and *SEM* of participants who were asked to judge aggression second (and trustworthiness first). Panel B shows participants' mean speed of judgments of aggression (white bars) versus judgments of trustworthiness (gray bars) as a function of whether they rated aggression first or second. * = Aggression judged faster than trustworthiness, $p < .05$. # = Judgments were faster if participants rated aggression first rather than second, $p < .05$. Panel C shows paired *t*-test values comparing the speed of aggression versus trustworthiness judgments within each observer. Positive values indicate that judgments of aggression were quicker, whereas negative values indicate judgments of trust were quicker. Dashed lines represent critical *t* values ($p < .05$, two-tailed).

Results from the bootstrapped mediation also indicated that the face ratio shared an indirect association with judgments of trustworthiness (95% CI = −3.37, −1.43). Using multilevel modeling, the average correlations between the face ratio and judgments of trustworthiness ($\gamma = -2.14$, $SE = 0.20$, $t[2,559] = -10.49$, $p < .001$), and between the face ratio and judgments of aggression ($\gamma = 2.72$, $SE = 0.28$, $t[2,559] = 9.68$, $p < .001$), were significant. The average association between judgments of aggression and trustworthiness was also significant ($\gamma = -0.34$, $SE = 0.04$, $t[2,559] = -9.06$, $p < .001$). The average relationship between the face ratio and trustworthiness was attenuated, but still significant ($\gamma = -1.29$, $SE = 0.25$, $t[2,558] = -5.22$, $p < .001$), when judgments of aggression ($\gamma = -0.31$, $SE = 0.04$, $t[2,558] = -7.95$, $p < .001$) was added as a simultaneous predictor, again indicating partial mediation (see Table 2 for Sobel test).[2]

Did the order in which participants provided ratings influence how they judged the faces? A mixed factorial ANOVA with one within-subjects factor (aggression vs. trustworthiness rating) and one between-subjects factor (aggression rated first vs. second) revealed a main effect of rating order ($F[1,128] = 282.70$, $p < .001$), such that those who rated aggression first rated faces as less aggressive and as less trustworthy (i.e., tended to use lower values on the 7-point rating scale for both ratings) than did those who rated aggression second. There was no main effect of rating type ($F[1,128] = 0.08$, $p = .78$) and no interaction between rating type and order of rating ($F[1,128] = 0.89$, $p = .35$). Therefore, although the group that rated aggression first tended to use lower values for both ratings on the 7-point scale than the group that rated aggression second, this change in rating order did not differentially influence ratings of aggression and trustworthiness. To determine whether this difference in the use of rating scales was indeed influenced by the order in which ratings were provided, or if the two groups had initial differences in the use of the 7-point scale, we analyzed the estimates of circle sizes made by participants before they provided judgments of aggression and trustworthiness. A paired samples *t* test indicated that the two groups significantly differed in their estimates of the size of the circles ($t[6] = 2.48$, $p = .05$), such that those who rated aggression first

[2] The results remained the same (partial mediation) when groups who rated aggression first versus second were analyzed separately.

rated circles as smaller (used lower values on the 7-point scale) than did those who rated aggression second (aggression first: $M = 4.61$, $SD = 2.05$; aggression second: $M = 4.80$, $SD = 1.90$). Therefore, as opposed to order of judgments influencing the way in which participants rated faces, it is more likely that the two groups differed at baseline in their use of the 7-point scales used for judgments.

We also found that the associations between the face ratio and judgments of aggression (aggression rated first: $r = .40$, $p < .01$; aggression rated second: $r = .55$, $p < .001$), and between the face ratio and judgments of trustworthiness (trustworthiness rated first: $r = -.36$, $p < .01$; trustworthiness rated second: $r = -.40$, $p < .01$), were significant irrespective of which rating was provided first. Fisher *r*-to-*z* transformations also confirmed that the strength of these associations did not differ between participants who rated aggression first versus second (aggression judgments: $z = -1.14$, $p = .25$; trustworthiness judgments: $z = 0.27$, $p = .78$). Ratings made by participants who judged aggression first were also highly correlated with the corresponding ratings made by those who judged aggression second (aggression: $r = .91$, $p < .001$; trustworthiness: $r = .83$, $p < .001$).

Therefore, the order with which participants provided ratings did not alter their judgments of aggression, of trustworthiness, or the relationship between both of these judgments and the face ratio.[3]

Do judgments of aggression share stronger links with the face ratio than do judgments of trustworthiness? To determine which judgment shared the strongest association with the face ratio, the associations between judgments of aggression and the face ratio, and between judgments of trustworthiness and the face ratio, were calculated for each observer across all four studies. The direction of the correlations between trustworthiness and the face ratio for each observer were reversed (multiplied by −1), so that their magnitude could be compared with those of aggression and the face ratio. These correlation coefficients were then transformed into Fisher *z* values and compared using paired samples *t* tests (see Figure 4). Across the studies, judgments of aggression shared significantly stronger associations with the face ratio than judgments of trustworthiness. A one sample *t* test was also used to test whether the differences in correlation magnitude across studies was significantly different from zero. This test again revealed that judgments of aggression shared significantly stronger associations with the face ratio across studies than did judgments of trustworthiness.

Discussion

Previous studies have identified links between judgments of aggression and trustworthiness (e.g., Carré et al., 2009; Oosterhof & Todorov, 2008), as well as links between each of these judgments and the face ratio (aggression and the face ratio: Boshyan et al., 2013; Carré et al., 2010; Geniole et al., 2012; Geniole & McCormick, 2013; Lefevre & Lewis, 2013; Short et al., 2012; trustworthiness and the face ratio: Efferson & Vogt, 2013; Kleisner et al., 2013; Stirrat & Perrett, 2010). The goal of the current studies was to determine the perceptual judgment to which the face ratio is most strongly and directly linked; such information would provide insight regarding the functional basis of the face ratio and the perceptual processes that shape judgments of aggression and trustworthiness. Based on the hypothesis that snap judgments of aggression may be more relevant for survival than snap judgments of trustworthiness, and that trustworthiness judgments may simply "piggyback" on judgments of aggression when exposure to a face is limited or reputational information is lacking, we predicted that (a) judgments of aggression would share stronger links with the face ratio than judgments of trustworthiness, (b) judgments of aggression would form faster than judgments of trustworthiness, and (c) judgments of aggression would mediate the link between the face ratio and judgments of trustworthiness. Across multiple sets of faces and using ratings provided by four different samples of observers, we found support for these predictions. Specifically, the face ratio shared a stronger relationship with judgments of aggression than with judgments of trustworthiness, judgments of aggression were made more quickly than were judgments of trustworthiness, and judgments of aggression mediated the face ratio-trustworthiness link, although there was evidence of only partial mediation when multilevel modeling was used. Thus, men with larger face ratios were judged as less trustworthy, in part, because they looked more aggressive than did men with smaller face ratios. These results also attest to the strength of the association between the face ratio and judgments of aggression; the relationship was found in four subsets of a new set of faces, bolstering our finding of a relationship in other studies (Boshyan et al., 2013; Carré et al., 2010; Geniole et al., 2012; Geniole & McCormick, 2013; Short et al., 2012).

We have proposed that the face ratio may have conferred adaptive benefits as a signal that readily conveys behavioral dispositions important to survival (Carré et al., 2009); judgments of aggression were associated with the face ratio even when photographs were shown for only 39 ms (Carré et al., 2010). Therefore, the face ratio can facilitate rapid assessments of formidability, which may serve to modulate speedy fight-or-flight responses when encountering a potentially dangerous stranger. Although judgments of trustworthiness are relevant for social interactions and can be formed after a short exposure time to a face as well

[3] To investigate whether the phrasing of the question about aggression in Study 4 ("How aggressive does this person look?") influenced participants' responses, we compared data from Study 4 with data from an ongoing study in which we used the same stimuli as in Study 4, but instead asked participants, "How aggressive would this person be if provoked?" Although this ongoing study also differed slightly in methodology (each photo was displayed for 1,000 ms, after which the aforementioned question appeared), it may provide some insight as to whether the phrasing of the question influences judgments of aggression. When we compare the responses from participants in Study 4 ($n = 40$) with those from participants in our ongoing study (current $n = 40$), correlations between the face ratio and mean judgments of aggression were not significantly different (Study 4, $r = .48$, $p < .001$; ongoing study, $r = .42$, $p < .001$; Fisher's *r*-to-*z* transformation to test difference between correlations: $z = 0.32$, $p = .75$) and the correlation between judgments of aggression made in Study 4 and those made in the ongoing study were high ($r = .94$, $p < .001$). An independent samples *t* test revealed that the ratings made by participants in Study 4 ($M = 3.56$, $SD = 0.87$) did not significantly differ ($t[128] = -0.81$, $p = .42$) from those made by participants in the ongoing study ($M = 3.67$, $SD = 0.77$). Thus, although we do not have experimental data directly testing whether the phrasing of the question about aggression alters responses, comparisons between data from Study 4 and data from our ongoing study suggest that this change in phrasing does not alter the judgments of aggression and the relationship these judgments of aggression share with the face ratio.

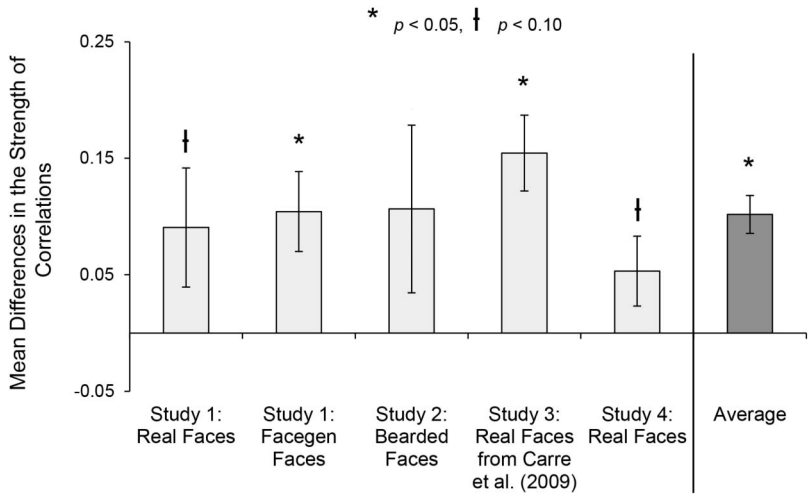

Figure 4. Bar graph showing differences in the strength of the associations between the face ratio and judgments of aggression compared with the strength of the associations between the face ratio and judgments of trustworthiness. Bars represent the mean difference in correlation strength such that values above the horizontal axis indicate stronger associations between the face ratio and judgments of aggression than between the face ratio and judgments of trustworthiness. The lighter bars show the mean difference within each study, whereas the dark bar shows the average difference across studies. Error bars represent standard errors of measurement.

(J. Willis & Todorov, 2006), the encounters in which judgments of trustworthiness are beneficial likely do not require such speed. Consequently, the face ratio, which can be gleaned rapidly from the face, may be more relevant for assessments of formidability and aggressive potential than for trustworthiness. Consistent with this possibility, we found that judgments of aggression were made more rapidly than judgments of trustworthiness.

Other researchers have reported that participants use aggressive adjectives more frequently than trustworthy adjectives when describing faces of strangers (Oosterhof & Todorov, 2008), indicating that aggression judgments may be more accessible than trustworthiness judgments. Further, based on a principal components analysis, the dimensions of valence and dominance accounted for 82% of the variability in judgment ratings (trustworthy, emotionally stable, responsible, sociable, caring, weird, attractive, mean, intelligent, aggressive, unhappy, confident, and dominant; Oosterhof & Todorov, 2008). Judgments of trustworthiness loaded only on the dimension of valence, whereas judgments of aggression loaded onto both dimensions. Thus, more relevant inferences about the characteristics within each dimension can be made from judgments of aggression than from judgments of trustworthiness. Further, participants reported higher confidence in judgments of aggression than in those of trustworthiness (see bottom panel of Figure 2 of J. Willis & Todorov, 2006). Thus, judgments of aggression may share stronger links with the face ratio because they are made more rapidly, are used more frequently to describe faces, are more relevant for making global inferences, and they are formed with greater confidence than are judgments of trustworthiness.

Our results for the speed with which judgments of aggression and trustworthiness are made were opposite to those reported by J. Willis and Todorov (2006). Specifically, the middle panel of Figure 2 in their article suggests that judgments of trustworthiness are made faster than are judgments of aggression. The discrepancy between their data and ours can likely be attributed to methodological differences. For example, we did not constrain the speed at which judgments could be made, whereas participants in Willis and Todorov could only provide a judgment after viewing photographs for durations of 100 to 1,000 ms; thus, they may have inadvertently prevented the recording of genuine, rapid responses.

Our results also provide insight as to why high-level cognitive judgments, such as those of trustworthiness, can form rapidly and with limited exposure to the face. Specifically, when there is little or no background information about an individual and exposure to the individual is brief, judgments such as trustworthiness may piggyback on, or extract information from, more primitive or survival-relevant social judgments, such as those of aggression, which are better suited for encounters with strangers. Thus, as opposed to judgments of aggression and trustworthiness sharing parallel perceptual processes such that facial features lead to the simultaneous formation of these judgments, it instead appears that these judgments are formed sequentially; facial features cue judgments of aggression, which, in turn, influence subsequent judgments of trustworthiness. Future studies may benefit from examining whether these results are generalizable to other social judgments that differ with regard to primacy or cognitive complexity (e.g., judgments of competence/intelligence vs. threat/danger/desirability).

It should be noted that although these mediation and speed-of-rating effects were relatively strong in the current study, judgments of aggression based on the face likely play less of a role in shaping judgments of trustworthiness after information about reputation or inferences from multiple interactions have been acquired. Indeed, when faces were manipulated using Facegen to appear less trustworthy rather than more trustworthy, participants invested less money in economic trust games (Rezlescu, Duchaine, Olivola, & Chater, 2012). Although this effect persisted when information about the other players' reputation was provided, it was attenuated

substantially. As such, in these circumstances, aggression (and similarly facial structure) may play less of a role in shaping judgments of trustworthiness.

Some researchers have proposed that the relationship between the face ratio and judgments of, or actual, behavior are based in learned social processes rather than natural selection. For example, different observers may be consistent in their ratings of faces because neutral expressions of certain faces may more or less resemble (albeit subtly) emotional expressions compared with those of other faces (emotional overgeneralization hypothesis, e.g., Montepare & Dobish, 2003; Said, Sebe, & Todorov, 2009). Angry expressions, for example, involve the lowering of the brow and the raising of the upper lip; both of which exaggerate the face ratio. Thus, men with larger face ratios may appear more aggressive when posed in a neutral expression than men with smaller face ratios because they look angrier. A recent study did report a positive association between the face ratio and the extent to which a face looked angry (Boshyan et al., 2013). Nevertheless, judgments of aggression were related to the face ratio when judgments of anger were controlled statistically, indicating that the face ratio guides judgments of aggressiveness over and above the extent to which it is associated with perceptions of anger. Additionally, although observers' judgments of aggression were sensitive to both how angry the faces of the men looked and to the size of the face ratio, the face ratio was related to the actual aggression of the men and anger was not. Thus, in a neutral face, the face ratio may be an "honest signal," whereas perceptions of anger may be misleading. Other research suggests that the decreased prosocial behavior among men with larger face ratios, compared with men with smaller face ratios, may be because of self-fulfilling prophecy; men with larger face ratios may act more antisocially because people anticipate, and thus elicit, such behavior through their own negative treatment of these men (Haselhuhn, Wong, & Ormiston, 2013). Nevertheless, how a bias arose to view wider faces as more aggressive is not readily explained by social learning.

Rapid detection of threat is imperative for survival, and detection of threat does not always require experience with either threat or faces. For example, rhesus monkeys reared in social isolation responded to facial displays of threat (Sackett, 1966). Thus, perceptual systems may have evolved to be highly sensitive to signals of threat. Further, the face ratio may be a signaling mechanism common to both human and nonhuman primates. A recent study found that capuchins with larger face ratios were more likely to achieve alpha status (Lefevre et al., 2014) and were more assertive (Wilson et al., 2014) than those with smaller face ratios. Nevertheless, whether nonhuman primates use this signal has yet to be investigated. We have found that judgments of aggression made by both children and adults are associated with the face ratio, even when the judgments are made of a face of an ethnicity for which the observer has little experience (e.g., ratings made in China of Caucasian faces, ratings made in Canada of Chinese faces; Short et al., 2012).

It is also likely that if perceptual sensitivity to a facial feature is adaptive, such sensitivity should be maintained despite the presence of facial hair, given our ancestral past likely involved social interactions with bearded rather than shaved men. In Study 2, judgments of aggression were associated with the face ratio even when faces had facial hair. Facial hair does not appear to cover the right and left zygion from which the bizygomatic width of the ratio is derived. Therefore, the face ratio may be a marker that is not obscured by facial hair. In contrast, other cues of masculinity and dominance (e.g., jaw line, chin size, lip size) may be less perceptible or more ambiguous when men are bearded. Facial hair likely evolved through intersexual and intrasexual selection processes, advertising a combination of traits such as aggressiveness, status, and reproductive potential (reviewed in Muscarella & Cunningham, 1996). Thus, our finding of a link between the face ratio and judgments of aggression, despite the presence of facial hair, provides further ecological validity of this association, and highlights the robust nature of the relationship.

There was an association between facial hair and the face ratio such that men with larger face ratios tended to have more facial hair. It may be that these features share an underlying endocrine mechanism. Testosterone secretion, for example, shapes the male face during puberty (Marečková et al., 2011; Verdonck, Gaethofs, Carels, & de Zegher, 1999) and also facilitates the growth of facial hair (e.g., Farthing, Mattei, Edwards, & Dawson, 1982). At the same time of development, testosterone pulses have organizational effects on brain regions involved in the regulation of aggression and other social behaviors (reviewed in Schulz, Molenda-Figueira, & Sisk, 2009), likely altering future responses to social interactions. Indeed, animal models suggest that pubertal androgens have long-lasting effects on aggressive behavior (e.g., Farrell & McGinnis, 2004). There is also some evidence in humans that testosterone concentrations measured during early adolescence are predictive of antisocial behavior measured years later (Drigotas & Udry, 1993). Thus, it is possible that both the face ratio and the amount of facial hair are markers of sensitivity to pubertal testosterone concentrations, which informs predictions about behavioral tendencies in adulthood. As such, sensitivity to both cues is likely advantageous when assessing formidability.

An unexpected finding was that the observers randomly assigned to the condition in which aggressiveness was rated first were faster at providing both judgments (aggression and trust) than were those assigned to the condition in which trustworthiness was rated first. This group difference was likely caused by these initial ratings, given that the two groups did not differ in response time to a general rating task (rating the size of circles) they completed before the facial judgments. Further, although ratings of aggression were faster than were ratings of trustworthiness in both conditions, ratings of aggression rated first were faster than ratings of aggression rated second, whereas ratings of trustworthiness rated first were slower than were ratings of trustworthiness when rated second. Thus, judgments of "aggression" facilitated subsequent judgments, whereas judgments of "trustworthiness" impeded subsequent judgments, perhaps because aggression requires speed whereas trust requires deliberation.

It should also be noted that irrespective of the ultimate mechanisms (e.g., rapid judgments of aggression may be more relevant for survival than rapid judgments of trustworthiness) that may account for the pattern of findings reported here, there are likely multiple proximate mechanisms involved. For example, the frequency with which the terms "aggressive" and "trustworthy" are used to describe faces may be the basis of the differences in the speed of the formation of these judgments; responding to a familiar term may be quicker than responding to an unfamiliar term. Consistent with this possibility, Oosterhof and Todorov (2008) reported that faces were more likely to be described as "aggres-

sive" and "mean" than as "trustworthy." In addition, Google's Ngram Viewer (as used in Greenfield, 2013; Michel et al., 2011) reported that, since 1985, the frequency of either the single word "aggressive" or the phrases "aggressive looking" or "aggressive face" was about 2 to 10 times more than that of the word "trustworthy," or the phrases "trustworthy looking" or "trustworthy face." It is well established that word frequency is negatively associated with RT (reviewed in Borowsky & Besner, 1993).

One limitation to these studies is that we did not include measures of the men's actual aggressive or untrustworthy behavior. Thus, we cannot determine whether judgments of aggression or of trustworthiness provided in the current study were accurate. In any case, despite variability in the extent to which social judgments are accurate (see Olivola & Todorov, 2010; Rule, Krendl, Ivcevic, & Ambady, 2013; Zebrowitz & Montepare, 2008), they nonetheless modulate behavior in social interactions (Rezlescu et al., 2012; van 't Wout & Sanfey, 2008) and decision making in many domains (reviewed in Olivola & Todorov, 2010). Therefore, the formation of social judgments is an important aspect of human psychology, independent of accuracy in such judgments. There are many examples in which facial features facilitate social judgments that are not always accurate (e.g., babyfacedness and intelligence or innocence, attractiveness and health; reviewed in Zebrowitz & Montepare, 2008). Further, some researchers have suggested that instead of focusing on possible errors in human judgments, it is worthwhile to examine the cognitive and perceptual systems that produce these judgments, as they likely served an adaptive purpose in the past (Haselton & Funder, 2006), thus leading to their maintenance and consistency in promoting social judgments. The association between the face ratio and judgments of aggression is consistent, and observers appear to be tuned to the face ratio across different age groups and cultures (Boshyan et al., 2013; Short et al., 2012). Further, irrespective of whether the face ratio accurately predicts the target's behavior, it influences the observer's behavior (e.g., Haselhuhn et al., 2013; Stirrat & Perrett, 2010), and thus is an important psychological or perceptual process worthy of investigation.

Conclusion

Judgments of trustworthiness are critical in regulating our social interactions and have been shown to modulate behavior in economic bargaining and trust games (Rezlescu, Duchaine, Olivola, & Chater, 2012; van 't Wout & Sanfey, 2008). Consistent with previous studies (e.g., Efferson & Vogt, 2013; Kleisner et al., 2013; Stirrat & Perrett, 2010), our results suggest that the face ratio is associated with judgments of trustworthiness. We extend these findings, however, and show that this link is not direct, but is instead mediated by judgments of aggression. Consistent with this mediation model, temporal precedence was also established: Judgments of aggression were provided faster by participants than were judgments of trustworthiness. Therefore, instead of judgments of aggression and trustworthiness forming simultaneously after the perception and processing of a face, it is more likely that judgments of aggression form first, prompting the subsequent formation of judgments of trustworthiness. As such, the formation of many social judgments, even those that are highly correlated, may be best characterized by a sequential "piggyback" framework, wherein cognitively complex judgments occur after, and extract information from, more primitive social judgments, especially when exposure to facial information is brief and reputational information is absent.

References

Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature, 393,* 470–474. doi:10.1038/30982

Alrajih, S., & Ward, J. (2013). Increased facial width-to-height ratio and perceived dominance in the faces of the UK's leading business leaders. Advance online publication. *British Journal of Psychology*. doi:10.1111/bjop.12035

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182. doi:10.1037/0022-3514.51.6.1173

Berry, D. S. (1990). Taking people at face value: Evidence for the kernel of truth hypothesis. *Social Cognition, 8,* 343–361. doi:10.1521/soco.1990.8.4.343

Blanchard, D. C., Griebel, G., Pobbe, R., & Blanchard, R. J. (2011). Risk assessment as an evolved threat detection and analysis process. *Neuroscience and Biobehavioral Reviews, 35,* 991–998. doi:10.1016/j.neubiorev.2010.10.016

Blanz, V., & Vetter, T. (1999, July). A morphable model for the synthesis of 3D faces. In *Proceedings of the 26th annual Conference on Computer Graphics and Interactive Techniques* (pp. 187–194). New York, NY: ACM Press/Addison Wesley.

Borowsky, R., & Besner, D. (1993). Word recognition: A multistage activation model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 813–840. doi:10.1037/0278-7393.19.4.813

Boshyan, J., Zebrowitz, L. A., Franklin, R. G., McCormick, C. M., & Carré, J. M. (2013). Age similarities in recognizing threat from faces and diagnostic cues. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences.* Advance online publication. doi:10.1093/geronb/gbt054

Carré, J. M., & McCormick, C. M. (2008). In your face: Facial metrics predict aggressive behaviour in the laboratory and in varsity and professional hockey players. *Proceedings. Biological Sciences/The Royal Society, 275,* 2651–2656. doi:10.1098/rspb.2008.0873

Carré, J. M., McCormick, C. M., & Mondloch, C. J. (2009). Facial structure is a reliable cue of aggressive behavior. *Psychological Science, 20,* 1194–1198. doi:10.1111/j.1467-9280.2009.02423.x

Carré, J. M., Morrissey, M. D., Mondloch, C. J., & McCormick, C. M. (2010). Estimating aggression from emotionally neutral faces: Which facial cues are diagnostic? *Perception, 39,* 356–377. doi:10.1068/pp.6543

Carré, J. M., Murphy, K. R., & Hariri, A. R. (2013). What lies beneath the face of aggression? *Social Cognitive and Affective Neuroscience, 8,* 224–229. doi:10.1093/scan/nsr096

Chambers, J. R., Swan, L. K., & Heesacker, M. (2014). Better off than we know: Distorted perceptions of incomes and income inequality in America. *Psychological Science, 25,* 613–618. doi:10.1177/0956797613504965

Christiansen, K., & Winkler, E. (1992). Hormonal, anthropometrical, and behavioral correlates of physical aggression in !Kung San men of Namibia. *Aggressive Behavior, 18,* 271–280. doi:10.1002/1098-2337(1992)18:4<271::AID-AB2480180403>3.0.CO;2-6

Coccaro, E. F., Sripada, C. S., Yanowitch, R. N., & Phan, K. L. (2011). Corticolimbic function in impulsive aggressive behavior. *Biological Psychiatry, 69,* 1153–1159. doi:10.1016/j.biopsych.2011.02.032

Deaner, R. O., Goetz, S. M. M., Shattuck, K., & Schnotala, T. (2012). Body weight, not facial width-to-height ratio, predicts aggression in pro

hockey players. *Journal of Research in Personality, 46,* 235–238. doi:10.1016/j.jrp.2012.01.005

de Souza, A. A. L., Baumgasten, V., Baiao, U., & Otta, E. (2003). Perception of men's personal qualities and prospect of employment as a function of facial hair. *Psychological Reports, 92,* 201–208. doi:10.2466/pr0.2003.92.1.201

Drigotas, S. M., & Udry, J. R. (1993). Biosocial models of adolescent problem behavior: Extension to panel design. *Social Biology, 40,* 1–7. doi:10.1080/19485565.1993.9988831

Efferson, C., & Vogt, S. (2013). Viewing men's faces does not lead to accurate predictions of trustworthiness. *Scientific Reports, 3,* 1047. doi:10.1038/srep01047

Erceg-Hurn, D. M., & Mirosevich, V. M. (2008). Modern robust statistical methods: An easy way to maximize the accuracy and power of your research. *American Psychologist, 63,* 591–601. doi:10.1037/0003-066X.63.7.591

Farrell, S. F., & McGinnis, M. Y. (2004). Long-term effects of pubertal anabolic–androgenic steroid exposure on reproductive and aggressive behaviors in male rats. *Hormones and Behavior, 46,* 193–203. doi:10.1016/j.yhbeh.2004.03.012

Farthing, M. J. G., Mattei, A. M., Edwards, C. R. W., & Dawson, A. M. (1982). Relationship between plasma testosterone and dihydrotestosterone concentrations and male facial hair growth. *British Journal of Dermatology, 107,* 559–564. doi:10.1111/j.1365-2133.1982.tb00406.x

Fink, B., Neave, N., & Seydel, H. (2007). Male facial appearance signals physical strength to women. *American Journal of Human Biology, 19,* 82–87. doi:10.1002/ajhb.20583

Fridlund, B. A. (1994). *Human facial expression: An evolutionary view*. San Diego, CA: Academic Press.

Geniole, S. N., Keyes, A. E., Carré, J. M., & McCormick, C. M. (2014). Fearless dominance mediates the relationship between the facial width-to-height ratio and willingness to cheat. *Personality and Individual Differences, 57,* 59–64. doi:10.1016/j.paid.2013.09.023

Geniole, S. N., Keyes, A. E., Mondloch, C. J., Carré, J. M., & McCormick, C. M. (2012). Facing aggression: Cues differ for female versus male faces. *PLoS ONE, 7,* e30366. doi:10.1371/journal.pone.0030366

Geniole, S. N., & McCormick, C. M. (2013). Taking control of aggression: Perceptions of aggression suppress the link between perceptions of facial masculinity and attractiveness. *Evolutionary Psychology, 11,* 1027–1043.

Goetz, S. M., Shattuck, K. S., Miller, R. M., Campbell, J. A., Lozoya, E., Weisfeld, G. E., & Carré, J. M. (2013). Social status moderates the relationship between facial structure and aggression. *Psychological Science, 24,* 2329–2334. doi:10.1177/0956797613493294

Gómez-Valdés, J., Hünemeier, T., Quinto-Sánchez, M., Paschetta, C., de Azevedo, S., González, M. F., . . . González-José, R. (2013). Lack of support for the association between facial shape and aggression: A reappraisal based on a worldwide population genetics perspective. *PLOS ONE, 8,* e52317. doi:10.1371/journal.pone.0052317

Greenfield, P. M. (2013). The changing psychology of culture from 1800 through 2000. *Psychological Science, 24,* 1722–1731. doi:10.1177/0956797613479387

Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate data analysis* (3rd ed.). New York, NY: Macmillan.

Haselhuhn, M. P., & Wong, E. M. (2012). Bad to the bone: Facial structure predicts unethical behaviour. *Proceedings. Biological Sciences/The Royal Society, 279,* 571–576. doi:10.1098/rspb.2011.1193

Haselhuhn, M. P., Wong, E. M., & Ormiston, M. E. (2013). Self-fulfilling prophecies as a link between men's facial width-to-height ratio and behavior. *PLoS ONE, 8,* e72259. doi:10.1371/journal.pone.0072259

Haselton, M. G., & Funder, D. C. (2006). The evolution of accuracy and bias in social judgment. In M. Schaller, J. A. Simpson, & D. T. Kenrick (Eds.), *Evolution and social psychology* (pp. 15–37). Philadelphia, PA: Psychology Press.

Hehman, E., Leitner, J. B., Deegan, M. P., & Gaertner, S. L. (2013). Facial structure is indicative of explicit support for prejudicial beliefs. *Psychological Science, 24,* 289–296. doi:10.1177/0956797612451467

Hehman, E., Leitner, J. B., & Gaertner, S. L. (2013). Enhancing static facial features increases intimidation. *Journal of Experimental Social Psychology, 49,* 747–754. doi:10.1016/j.jesp.2013.02.015

Kenny, C. T., & Fletcher, D. (1973). Effects of beardedness on person perception. *Perceptual and Motor Skills, 37,* 413–414. doi:10.2466/pms.1973.37.2.413

Kleisner, K., Priplatova, L., Frost, P., & Flegr, J. (2013). Trustworthy-looking face meets brown eyes. *PLoS ONE, 8,* e53285. doi:10.1371/journal.pone.0053285

Kramer, R. S. S., King, J. E., & Ward, R. (2011). Identifying personality from the static, nonexpressive face in humans and chimpanzees: Evidence of a shared system for signalling personality. *Evolution and Human Behavior, 32,* 179–185. doi:10.1016/j.evolhumbehav.2010.10.005

Kramer, R. S. S., & Ward, R. (2012). Cues to personality and health in the facial appearance of chimpanzees (*Pan troglodytes*). *Evolutionary Psychology, 10,* 320–337.

Lai, M. C., Lombardo, M. V., Ruigrok, A. N., Chakrabarti, B., Wheelwright, S. J., Auyeung, B., . . . Baron-Cohen, S. (2012). Cognition in males and females with autism: Similarities and differences. *PLoS ONE, 7,* e47198. doi:10.1371/journal.pone.0047198

Lefevre, C. E., & Lewis, G. J. (2013). Perceiving aggression from facial structure: Further evidence for a positive association with facial width-to-height ratio and masculinity, but not for moderation by self-reported dominance. *European Journal of Personality*. Advance online publication. doi:10.1002/per.1942

Lefevre, C. E., Wilson, V. A. D., Morton, F. B., Brosnan, S., Paukner, A., & Bates, T. C. (2014). Facial width-to-height ratio relates to alpha status and assertive personality in capuchin monkeys. *PLoS ONE, 9,* e93369. doi:10.1371/journal.pone.0093369

Marečková, K., Weinbrand, Z., Chakravarty, M. M., Lawrence, C., Aleong, R., Leonard, G., . . . Paus, T. (2011). Testosterone-mediated sex differences in the face shape during adolescence: Subjective impressions and objective features. *Hormones and Behavior, 60,* 681–690. doi:10.1016/j.yhbeh.2011.09.004

Mason, R. L., Gunst, R. F., & Hess, J. L. (1989). *Statistical design and analysis of experiments: Applications to engineering and science*. New York, NY: Wiley

McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review, 90,* 215–238. doi:10.1037/0033-295X.90.3.215

McCormick, C. M. (2013). Watch where and how you stick pins when playing with voodoo correlations. *Journal of General Psychology, 140,* 82–86. doi:10.1080/00221309.2012.737872

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., & Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books, *Science, 331,* 176–182. doi:10.1126/science.1199644

Montepare, J. M., & Dobish, H. (2003). The contribution of emotion perceptions and their overgeneralizations to trait impressions. *Journal of Nonverbal Behavior, 27,* 237–254. doi:10.1023/A:1027332800296

Mueller, E. M., Makeig, S., Stemmler, G., Hennig, J., & Wacker, J. (2011). Dopamine effects on human error processing depend on catechol-O-methyltransferase VAL158MET genotype. *The Journal of Neuroscience, 31,* 15818–15825. doi:10.1523/JNEUROSCI.2103-11.2011

Muscarella, F., & Cunningham, M. R. (1996). The evolutionary significance and social perception of male pattern baldness and facial hair. *Ethology & Sociobiology, 17,* 99–117. doi:10.1016/0162-3095(95)00130-1

Neter, J., Wasserman, W., & Kutner, M. H. (1989). *Applied linear regression models*. Homewood, IL: Irwin.

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. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 105,* 11087–11092. doi:10.1073/pnas.0805664105

Özener, B. (2012). Facial width-to-height ratio in a Turkish population is not sexually dimorphic and is unrelated to aggressive behavior. *Evolution and Human Behavior, 33,* 169–173. doi:10.1016/j.evolhumbehav.2011.08.001

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40,* 879–891. doi:10.3758/BRM.40.3.879

Preacher, K. J., & Leonardelli, G. J. (2010). Calculation for the Sobel test: An interactive calculation tool for mediation tests. Retrieved from http://quantpsy.org/sobel/sobel.htm

Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2004). *HLM 7 for Windows* [Computer software]. Lincolnwood, IL: Scientific Software International.

Rezlescu, C., Duchaine, B., Olivola, C. Y., & Chater, N. (2012). Unfakeable facial configurations affect strategic choices in trust games with or without information about past behavior. *PLoS ONE, 7,* e34293. doi:10.1371/journal.pone.0034293

Rule, N. O., Krendl, A. C., Ivcevic, Z., & Ambady, N. (2013). Accuracy and consensus in judgments of trustworthiness from faces: Behavioral and neural correlates. *Journal of Personality and Social Psychology, 104,* 409–426. doi:10.1037/a0031050

Ruppert, D. (1988). Trimming and Winsorization. In S. Kotz, N. L. Johnson, & C. B. Read (Eds.), *Encyclopedia of Statistical Sciences,* Vol. 9, (pp. 348–353). New York: Wiley. doi:10.1002/0471667196.ess2768.pub2

Sackett, G. P. (1966). Monkeys reared in isolation with pictures as visual input: Evidence for an innate releasing mechanism. *Science, 154,* 1468–1473. doi:10.1126/science.154.3755.1468

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

Schulz, K. M., Molenda-Figueira, H. A., & Sisk, C. L. (2009). Back to the future: The organizational–activational hypothesis adapted to puberty and adolescence. *Hormones and Behavior, 55,* 597–604. doi:10.1016/j.yhbeh.2009.03.010

Sell, A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., & Gurven, M. (2009). Human adaptations for the visual assessment of strength and fighting ability from the body and face. *Proceedings. Biological Sciences/The Royal Society, 276,* 575–584. doi:10.1098/rspb.2008.1177

Short, L. A., Mondloch, C. J., McCormick, C. M., Carré, J. M., Ma, R., Fu, G., & Lee, K. (2012). Detection of propensity for aggression based on facial structure irrespective of face race. *Evolution and Human Behavior, 33,* 121–129. doi:10.1016/j.evolhumbehav.2011.07.002

Singular Inversions. (2010). Facegen modeller, Version 3.5 [Computer software]. Retrieved from http://www.facegen.com

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology, 13,* 290–312. doi:10.2307/270723

Stephen, I. D., Oldham, F. H., Perrett, D. I., & Barton, R. A. (2012). Redness enhances perceived aggression, dominance and attractiveness in men's faces. *Evolutionary Psychology, 10,* 562–572.

Stillman, T. F., Maner, J. K., & Baumeister, R. F. (2010). A thin slice of violence: Distinguishing violent from nonviolent sex offenders at a glance. *Evolution and Human Behavior, 31,* 298–303. doi:10.1016/j.evolhumbehav.2009.12.001

Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust. *Psychological Science, 21,* 349–354. doi:10.1177/0956797610362647

Stirrat, M., Stulp, G., & Pollet, T. V. (2012). Male facial width is associated with death by contact violence: Narrow-faced males are more likely to die from contact violence. *Evolution and Human Behavior, 33,* 551–556. doi:10.1016/j.evolhumbehav.2012.02.002

Taylor, P. W., & Elwood, R. W. (2003). The mismeasure of animal contests. *Animal Behaviour, 65,* 1195–1202. doi:10.1006/anbe.2003.2169

Todorov, A., Mende-Siedlecki, P., & Dotsch, R. (2013). Social judgments from faces. *Current Opinion in Neurobiology, 23,* 373–380. doi:10.1016/j.conb.2012.12.010

Townsend, S. S., Eliezer, D., Major, B., & Mendes, W. B. (2014). Influencing the world versus adjusting to constraints social class moderates responses to discrimination. *Social Psychological and Personality Science, 5,* 226–234. doi:10.1177/1948550613490968

Třebický, V., Havlíček, J., Roberts, S. C., Little, A. C., & Kleisner, K. (2013). Perceived aggressiveness predicts fighting performance in mixed-martial-arts fighters. *Psychological Science, 24,* 1664–1672. doi:10.1177/0956797613477117

van 't Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition, 108,* 796–803. doi:10.1016/j.cognition.2008.07.002

Verdonck, A., Gaethofs, M., Carels, C., & de Zegher, F. (1999). Effect of low-dose testosterone treatment on craniofacial growth in boys with delayed puberty. *European Journal of Orthodontics, 21,* 137–143. doi:10.1093/ejo/21.2.137

Weston, E. M., Friday, A. E., & Liò, P. (2007). Biometric evidence that sexual selection has shaped the hominin face. *PLoS ONE, 2,* e710>. doi:10.1371/journal.pone.0000710

Wilcox, R. R. (2005). Outlier detection. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science,* Vol. 3 (pp. 1494–1497). Chichester, UK: John Wiley & Sons. doi:10.1002/0470013192.bsa742

Wilkowski, B. M., & Meier, B. P. (2010). Bring it on: Angry facial expressions potentiate approach-motivated motor behavior. *Journal of Personality and Social Psychology, 98,* 201–210. doi:10.1037/a0017992

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17,* 592–598. doi:10.1111/j.1467-9280.2006.01750.x

Willis, M. L., Dodd, H. F., & Palermo, R. (2013). The relationship between anxiety and the social judgements of approachability and trustworthiness. *PLoS ONE, 8,* e76825. doi:10.1371/journal.pone.0076825

Wilson, V., Lefevre, C. E., Morton, F. B., Brosnan, S., Paukner, A., & Bates, T. C. (2014). Personality and facial morphology: Links to assertiveness and neuroticism in capuchins (Sapajus [*Cebus*] apella). *Personality and Individual Differences, 58,* 89–94. doi:10.1016/j.paid.2013.10.008

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass, 2,* 1497–1517. doi:10.1111/j.1751-9004.2008.00109.x

Received November 18, 2013
Revision received March 24, 2014
Accepted March 24, 2014 ■