Perception, 2010, volume 39, pages 356 – 377

doi:10.1068/p6543

Estimating aggression from emotionally neutral faces: Which facial cues are diagnostic?

Justin M Carré, Mark D Morrissey, Catherine J Mondloch§, Cheryl M McCormick¶§
Department of Psychology (§ also Centre for Neurosciences), Brock University, St Catharines, Ontario L2S 3A1, Canada; e-mail: cmccormick@brocku.ca
Received 5 August 2009, in revised form 21 December 2009

Abstract. The facial width-to-height ratio, a size-independent sexually dimorphic property of the human face, is correlated with aggressive behaviour in men. Furthermore, observers' estimates of aggression from emotionally neutral faces are accurate and are highly correlated with the facial width-to-height ratio. We investigated whether observers use the facial width-to-height ratio to estimate propensity for aggression. In experiments 1a – 1c, estimates of aggression remained accurate when faces were blurred or cropped, manipulations that reduce featural cues but maintain the facial width-to-height ratio. Accuracy decreased when faces were scrambled, a manipulation that retains featural information but disrupts the facial width-to-height ratio. In experiment 2, computer-modeling software identified eight facial metrics that correlated with estimates of aggression; regression analyses revealed that the facial width-to-height ratio was the only metric that uniquely predicted these estimates. In experiment 3, we used a computer-generated set of faces varying in perceived threat (Oosterhof and Todorov, 2008 *Proceedings of the National Academy of Sciences of the USA* **105** 11087 – 11092) and found that as emotionally neutral faces became more 'threatening', the facial width-to-height ratio increased. Together, these experiments suggest that the facial width-to-height ratio is an honest signal of propensity for aggressive behaviour.

1 Introduction

Despite the adage "don't judge a book by its cover", most people believe they can accurately judge someone's character from static facial cues (Hassin and Trope 2000). Consensus in trait impressions can be achieved with minimal exposure to faces (eg within 100 ms—Bar et al 2006; Willis and Todorov 2006; Rule and Ambady 2008a; Carré et al 2009) and, for some traits, these judgments are accurate (eg extraversion and conscientiousness—Penton-Voak et al 2006; Little and Perrett 2007; sociosexuality—Boothroyd et al 2008; Stillman and Maner 2009; interest in children—Roney et al 2006; sexual orientation—Rule and Ambady 2008a; Rule et al 2008; fighting ability—Sell et al 2009; and propensity for aggression—Carré et al 2009). Furthermore, judgments based on the face are known to predict various social outcomes, including election results (Todorov et al 2005; Ballew and Todorov 2007; Little et al 2007), financial success (Rule and Ambady 2008b), military rank (Mazur et al 1984), and sentencing decisions (Blair et al 2004).

Some of the facial cues that influence trait judgments have been identified. For example, faces with large eyes, high eyebrows, small chins, and low vertical placement of facial features are rated as being more 'babyfaced' (Zebrowitz 1997). However, there is little support for the accuracy of estimates based on 'babyfacedness' (see Zebrowitz and Montepare 2008, for review). We recently reported that the facial width-to-height ratio, a size-independent sexually dimorphic property of the human face (Weston et al 2007), is correlated with aggressive behaviour among men tested in the laboratory and among varsity and professional hockey players during competition (Carré and McCormick 2008a). An earlier investigation of aggression among the Kung San hunter-gatherer tribe from Namibia found that men with a violent history had wider faces

¶ Author to whom all correspondence should be addressed.

(bizygomatic width, a component of the facial width-to-height ratio) than men without a violent history (Christiansen and Winkler 1992). Further, observers were accurate in estimating propensity for aggression on viewing photographs of the men whose aggressive behaviour was measured in the lab, even when exposure to the stimulus faces was limited to 39 ms (Carré et al 2009). Judgments of aggression were highly consistent across observers and between men and women (no sex differences in ratings of aggression in the facial stimuli). In addition, the observers' estimates of aggression were highly associated with the facial width-to-height ratio of the stimulus faces. Thus, the facial width-to-height ratio may serve as an 'honest signal' of propensity for aggression that may be used to guide adaptive social behaviour (Carré et al 2009). However, it remains possible that estimates of aggression from emotionally neutral faces are based on other facial metrics that may or may not be correlated with the facial width-to-height ratio.

In the following experiments, we used three approaches to investigate more fully whether the facial width-to-height ratio is the basis for estimates of aggression or whether other cues in the face convey the critical information. The first question was whether perceivers can accurately predict aggression when the facial width-to-height ratio is preserved but other facial cues (eg contour, facial features) are eliminated or their salience is reduced and, conversely, when other facial cues remain intact but configural cues, including the facial width-to-height ratio, are eliminated. In experiments 1a through 1c, the facial stimuli were modified to eliminate facial contour (by cropping the images), featural cues (by blurring the images), or the facial width-to-height ratio (by scrambling the images). The second question was to what extent other facial cues are diagnostic of aggression. In experiment 2, 61 metrics other than the facial width-to-height ratio in the unmanipulated facial stimuli were quantified with the algorithm provided by Facegen software; we then examined whether any of these 61 metrics were associated with estimates of aggression. The third question was the extent to which faces manipulated to be perceived as more/less threatening vary in facial width-to-height ratio. In experiment 3 we analysed a set of faces generated by Oosterhof and Todorov (2008) using the Facegen software in their investigations of cues in the face that convey threat. Such a correspondence would be consistent with our hypothesis that perceivers can use facial width-to-height ratio to predict threat, a trait that is conceptually related to aggression.

2 Experiment 1. Modifications of facial stimuli to eliminate features or facial width-to-height ratio

2.1 *Method*

2.1.1 *Participants.* All participants were undergraduate students who received either a Can \$5 honorarium or partial course credit for participation. Experiment 1a involved eight women and eight men (mean age = 23.81 years, SD = 6.18 years); experiment 1b involved sixteen women (mean age = 19.38 years, SD = 1.41 years); experiment 1c involved eighteen women and two men (mean age = 23.85 years, SD = 4.77 years); and experiment 1d involved sixteen women (mean age = 19.25 years, SD = 0.97 years). Thus, experiment 1 consisted of four sub-experiments. All procedures were approved by the university's Research Ethics Board.

2.1.2 *Stimuli.* Photographs were obtained from a sample of 37 men for which aggressive behaviour and facial width-to-height ratio was quantified previously (for more detail, see Carré and McCormick 2008a). In brief, these men were volunteers from an Introductory Psychology participant pool who received a Can \$5 honorarium and partial course credit for their participation. Aggressive behaviour was measured with a modified version of the Point Subtraction Aggression Paradigm, a well-validated laboratory task (eg Cherek et al 1996). This task was designed to assess aggressive behaviour in response to provocation (ie reactive aggression). In brief (see Carré and McCormick 2008b for full description), participants were led to believe that they were playing a computer game with another individual who was located in an adjacent room. Participants were instructed that the goal of the game was to earn as many points as possible, and that they would be compensated on the basis of the number of total points earned during the task. Participants had three response options available to them: option 1 = point reward, option 2 = point steal, and option 3 = point protection. Hitting option 1 a hundred consecutive times caused their point-counter to enlarge, flash several times with positive signs, and increase by one point, indicating that they earned a point. Participants were instructed that, throughout the task, it may occur that their point-counter turns red, flashes several times with negative signs, and decreases by one point, indicating that the person paired with them stole a point. Participants were told the points stolen from them would be added to their partner's point-counter. In addition to option 1 (point reward), participants could also choose to select option 2 (point steal) or option 3 (point protection). Hitting option 2 ten consecutive times would result in their partner losing 1 point. However, participants were notified that, although their partner would lose a point, they had been randomly assigned to the experimental condition whereby the points that they stole would not be added to their point-counter. Hitting option 3 ten consecutive times protected participants' point-counter against point subtractions (ie steal) for a variable amount of time.

Given that participants did not gain any financial reward from stealing, and that stealing occurred in response to provocation, it can be inferred that stealing points served to 'punish' one's partner, and as such represents our primary measure of aggressive behaviour. Our use of the term aggression in this context is consistent with the widely held definition of aggression as "any form of behavior directed toward the goal of harming or injuring another living being who is motivated to avoid such treatment" (Baron and Richardson 1994, page 7). Importantly, the harm or injury does not need to be physical in nature, but simply needs to be considered an aversive stimulus by the receiver.

The facial width-to-height ratio of the men was measured with NIH ImageJ software and involved the landmarks originally used by Weston and colleagues (2007). Specifically, the distance between the left and right zygion (bizygomatic width) was divided by the distance between the upper lip and mid-brow (upper facial height) to yield the facial width-to-height ratio. For the present experiments, as in our previous studies (Carré et al 2009), the sample of stimulus faces was reduced to include only Caucasian men without facial hair (to avoid judgments based on stereotypes) and displaying neutral expressions ($n = 24$, mean age = 19.08 years, SD = 1.41 years). There was no bias in the selection of stimuli; in fact, the relationship between actual aggression and facial width-to-height ratio is slightly reduced by the reduction in sample size (from $r = 0.38$ with $n = 37$ total sample, to $r = 0.31$ in the reduced sample), thereby providing a conservative test of our hypothesis. Faces were converted to 8-bit grey-scale and were standardised by using a hairline–chin distance of 400 pixels and placed within a black background.

Each original face (see figure 1a) was modified with Adobe Photoshop (version 6) to create four new conditions. In the 'chin/forehead crop' condition (see figure 1b), forehead and lower jaw were removed, but the facial width-to-height ratio was maintained. In the 'side crop' condition (see figure 1c), the width of the faces in 'chin/forehead crop' was cropped again to exclude vertical information beyond the outer eyebrows (eg ears and mid-jaw line), while maintaining the same rank order of the facial width-to-height ratio as in the original faces and of the 'chin/forehead crop' (ie facial width-to-height ratios from 'side crop' were strongly correlated with facial width-to-height ratios from 'chin/forehead crop', $r = 0.95$). These two conditions allowed us to examine whether any influence of

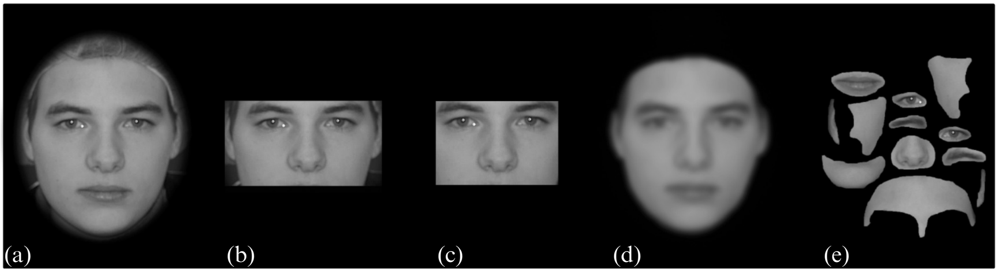

Figure 1. Example of faces used in experiment 1: (a) unmanipulated, (b) chin/forehead crop, (c) side crop, (d) blur, and (e) scramble.

the facial width-to-height ratio would be maintained in the absence of the external contour. 'Blur' and 'scramble' face conditions (see figures 1d and 1e, respectively) were based on the procedures in Hayward and colleagues (2008). 'Blur' faces were created with a Gaussian filter with a radius of 3 pixels in Adobe Photoshop (as in Collishaw and Hole 2000). 'Blur' faces retain configural information, including the facial width-to-height ratio, but disrupt featural processing, so much so that, when faces are both blurred and scrambled, recognition of facial identity drops to chance levels (Collishaw and Hole 2000; Hayward et al 2008). 'Scramble' faces, designed to eliminate all configural cues including the facial width-to-height ratio, were created by cutting each face into 12 pieces and placing the pieces into one of four arrangements. Blurred faces allowed us to examine the influence of the facial width-to-height ratio when potential featural cues were reduced in salience; scrambled faces allowed us to examine the influence of individual features in the absence of a face context.

2.2 *Experiment 1a*

2.2.1 *Procedure.* Stimulus faces were presented on a black background with E-Prime software on a 14-inch LCD monitor and were approximately 17 cm wide by 20 cm high (or 15.2 deg by 12.9 deg when viewed from 75 cm). Participants were told how the aggressive behaviour of the men had been assessed. The unmanipulated facial stimuli were presented in random order for 1000 ms to familiarise the participants with the range of faces. Next, each face within each stimulus condition was presented for 2000 ms (fully randomised), after which the question "How aggressive would this person be if provoked?" appeared on a black background along with a seven-point Likert scale (1—not at all aggressive, to 7—very aggressive). Participants were given unlimited time to make their response on a numerical keypad, which then caused the next stimulus face to appear. Within each stimulus condition, faces were presented in a random order and the order of conditions was semi-counterbalanced. Specifically, 'chin/forehead crop' and 'side crop' were always presented in blocks 1 or 2 with their order counterbalanced, and 'blur' and 'scramble' were always presented in blocks 3 or 4 with their order counterbalanced.

2.2.2 *Results.* Cronbach's α was calculated to examine the consistency of the ratings of estimated aggression across individual participants. The estimates of aggression were highly consistent across the sixteen individual observers for the 'chin/forehead crop', 'side crop', and 'blur' conditions, but not for the 'scramble' condition (Cronbach's $\alpha = 0.87$, 0.83, 0.92, and 0.57, respectively). For each participant we calculated the correlation between his/her estimate of aggression for the 24 faces and both the facial width-to-height ratio and actual aggression of the stimulus faces. One-sample t-tests were computed to test the primary hypothesis that these correlations would be significantly different from the null hypothesis (ie no association) in all conditions except the 'scramble' condition.

The t-tests indicated that the correlations between the estimates of aggression by individual participants and the facial width-to-height ratios of the targets were positive in all conditions (ps < 0.001 for 'chin/forehead crop', 'side crop', and 'blur'; $p = 0.05$ for 'scramble') and with actual aggression in all conditions (ps < 0.005) (see right panels of figure 2 and figure 3, respectively). The 95% confidence intervals for the relationship between estimated aggression and facial width-to-height ratio and estimated aggression and actual aggression approached zero (ie between 0.0005 and 0.055) in only one of the four conditions: 'scramble' faces. We also examined the individual components of the facial width-to-height ratio: width (distance between left and right zygion) and height (distance between mid brow and upper lip), in relation to estimates of aggression. Across all four conditions ('chin/forehead crop', 'side crop', 'blur', and 'scramble'), the facial width of the targets was positively correlated with observer estimates of aggression (all ps < 0.01). For 'chin/forehead crop' and 'side crop', upper facial height was negatively correlated with observer estimates of aggression (ps < 0.02).

Effect sizes (Cohen's d) were calculated to examine the strength of the relationship between the observers' estimates of aggression and either the facial width-to-height ratio, facial width, or facial height in each of the four conditions. Effect sizes indicated that the facial width-to-height ratio was more strongly associated with observer estimates of aggression than either the width or height alone. For the relationship between estimates of aggression and facial width-to-height ratio, Cohen's $d = 3.29$ for 'chin/forehead crop', 2.40 for 'side crop', 2.23 for 'blur', and 0.53 for 'scramble'. For the relationship between estimates of aggression and width, Cohen's $d = 0.81$, 0.92, 1.78, and 0.84, respectively. For the relationship between estimates of aggression and height, Cohen's $d = 1.17$, 0.68, 0.36, and 0.29, respectively.

To investigate the effect of stimulus type, the above correlations were converted to Fisher's z-scores, and then repeated-measures ANOVA were computed to examine whether the association between estimated aggression and facial width-to-height ratio or estimated aggression and actual aggression differed across experimental conditions. Stimulus condition was a factor in the relationship between estimated aggression and facial width-to-height ratio ($F_{3,45} = 11.24$, $p < 0.001$). Planned comparison indicated that the relationship was greater for 'chin/forehead crop' than for 'scramble' ($t_{15} = 4.46$, $p < 0.001$), but did not differ between 'chin/forehead crop' and 'side crop' ($t_{15} = 0.14$, $p = 0.89$) or 'chin/forehead crop' and 'blur' ($t_{15} = 0.21$, $p = 0.84$). The relationship between estimated aggression and actual aggression did not differ as a function of stimulus type, although the effect approached statistical significance ($F_{3,45} = 2.37$, $p = 0.08$) with the relationship between estimated aggression and actual aggression lowest for the 'scramble' condition.

For each stimulus face, we calculated the mean estimated aggression across participants and correlated that with both the facial width-to-height ratio and actual aggression of the stimulus faces. The mean estimated aggression of each stimulus face across participants was significantly correlated with the facial width-to-height ratio and with actual aggression in the 'chin/forehead crop', 'side crop', and 'blur' conditions (all ps < 0.05), but not in the 'scramble' condition (ps $= 0.07$ and 0.08, respectively) (see left panels of figures 2 and 3, respectively).

2.3 *Experiment 1b*

2.3.1 *Procedure.* In experiment 1a, the 'blur' and 'scramble' conditions always occurred either third or fourth, after the 'chin/forehead crop' and 'side crop' and after a pre-test exposure to the unmanipulated faces, thus providing significant opportunity for learning to have occurred. In this experiment we investigated whether estimates of aggression in the 'blur' and 'scramble' conditions would remain accurate when the possibility of learning from exposure to the unmanipulated faces was reduced and when the stimulus

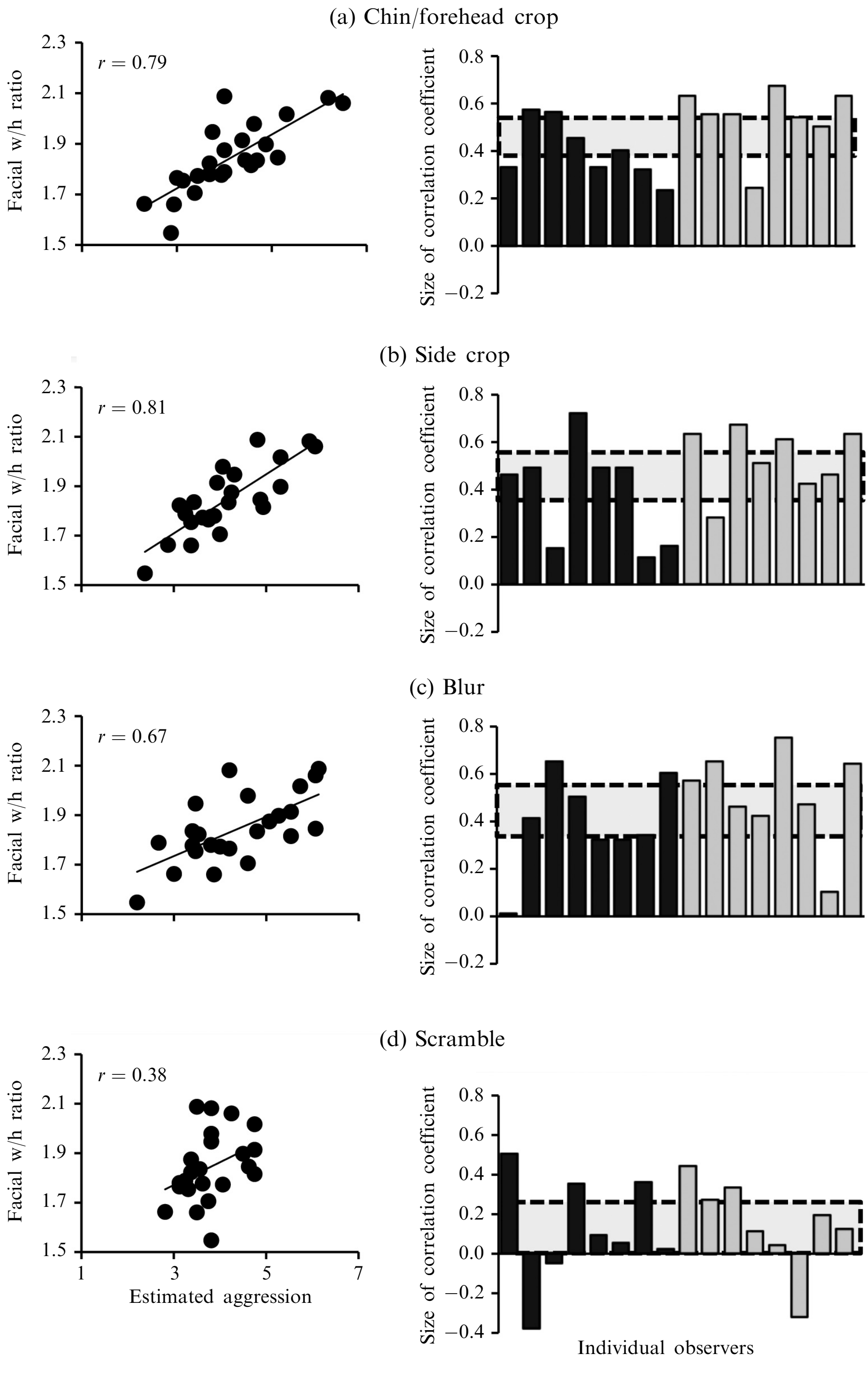

Figure 2. Left panel: relationship between mean estimated aggression across participants and facial width-to-height ratio (facial w/h ratio) for (a) chin/forehead crop, (b) side crop, (c) blur, and (d) scramble. Right panel: relationship for each individual observer between estimated aggression and facial width-to-height ratio (facial w/h ratio) for (a) chin/forehead crop, (b) side crop, (c) blur, and (d) scramble. Black bars: female observers ($n = 8$); grey bars: male observers ($n = 8$). Shaded areas represent the 95% confidence intervals (CI). Mean r-value of (a) $= 0.47$ (95% CI $= 0.39-0.55$); mean r-value of (b) $= 0.46$ (95% CI $= 0.35-0.56$); mean r-value of (c) $= 0.45$ (95% CI $= 0.34-0.56$); mean r-value of (d) $= 0.13$ (95% CI $= 0.0005-0.26$).

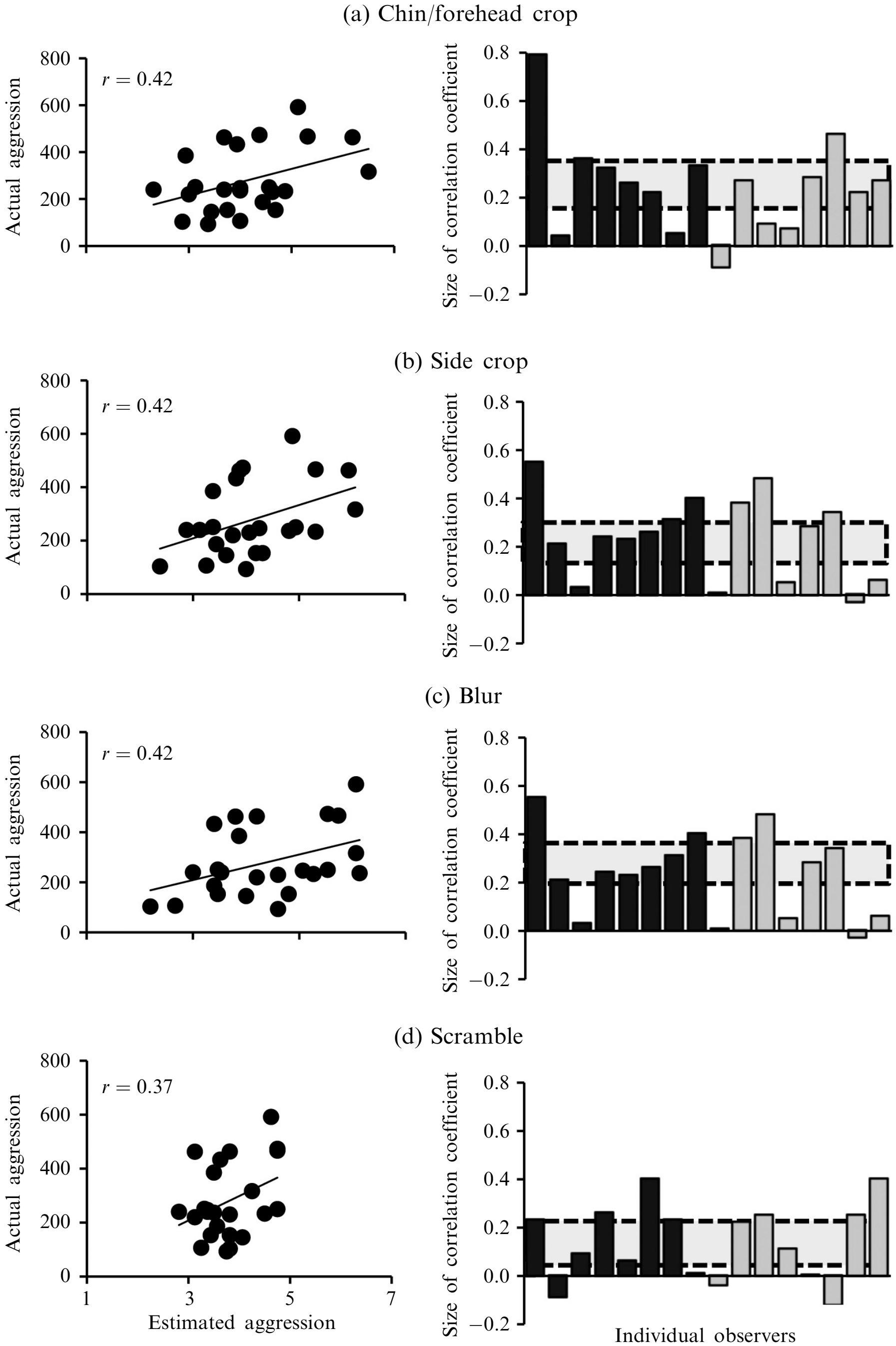

Figure 3. Left panel: relationship between mean estimated aggression across participants and actual aggression for (a) chin/forehead crop, (b) side crop, (c) blur, and (d) scramble. Right panel: relationship for each individual observer between estimated aggression and actual aggression for (a) chin/forehead crop, (b) side crop, (c) blur, and (d) scramble. Black bars: female observers ($n = 8$); grey bars: male observers ($n = 8$). Shaded areas represent the 95% confidence intervals (CI). Mean r-value of (a) = 0.25 (95% CI = 0.14 – 0.36); mean r-value of (b) = 0.24 (95% CI = 0.14 – 0.33); mean r-value of (c) = 0.29 (95% CI = 0.21 – 0.37); mean r-value of (d) = 0.14 (95% CI = 0.06 – 0.23).

exposure time for each condition was reduced from 2000 ms to 39 ms. The three stimulus conditions used were the unmanipulated faces, and the 'blur' and 'scramble' faces as in experiment 1a. Each of the 24 faces was presented in each condition for 39 ms only. The unmanipulated condition was always presented first, and the order of the 'blur' and 'scramble' conditions was counterbalanced. The order of faces was fully randomised within each condition. The presentation of a stimulus face was preceded by a central fixation cross that appeared for 500 ms. A face was then presented for 39 ms, after which the question "How aggressive would this person be if provoked?" appeared on a black background along with a seven-point Likert scale (1—not at all aggressive, to 7—very aggressive). As in experiment 1a, participants were told how the aggressive behaviour of the men had been assessed before the presentation of stimulus faces. In addition, after completion of the experiment, participants were asked "What facial cues did you use to make your judgements?", and responses were recorded by the experimenter.

2.3.2 *Results.* The estimates of aggression were relatively consistent across individual observers for the 'blur' condition (Cronbach's $\alpha = 0.68$) but not for the 'scramble' condition (Cronbach's $\alpha = 0.36$). For each participant we calculated the correlation between his/her estimate of aggression for the 24 faces and both the facial width-to-height ratio and actual aggression of the stimulus faces. The results for the unmanipulated condition (always presented first) were reported previously and are not included here, except to note that significant associations between estimates of aggression and actual aggression and estimates of aggression and the facial width-to-height ratio were found for the unmanipulated faces (Carré et al 2009). One-sample t-tests were computed to test the primary hypothesis that these correlations would be significantly different from the null hypothesis (ie no association) in the 'blur' condition, but not in the 'scramble' condition. The t-tests indicated that the estimates of aggression of individual participants were positively correlated with facial width-to-height ratio and actual aggression in the 'blur' condition ($t_{15} = 5.07, p < 0.001$; and $t_{15} = 3.88, p < 0.001$, respectively—see right panel of figure 4a), but not the 'scramble' condition ($t_{15} = -0.29$, $p = 0.78$; and $t_{15} = -1.33$, $p = 0.20$, respectively—see right panel of figure 4b). For 'scramble' faces, the 95% confidence intervals for the relationship between estimated aggression and facial width-to-height ratio and estimated aggression and actual aggression included zero. To compare the two experimental conditions, the correlations were converted to Fisher's z-scores, and then we computed paired-sample t-tests on the Fisher's z transformed conditions. The paired-sample t-tests indicated that the correlations between estimates of aggression and the facial width-to-height ratio and the correlations between estimates of aggression and actual aggression were significantly higher in the 'blur' condition than in the 'scramble' condition ($t_{15} = 3.81$, $p = 0.002$; and $t_{15} = 3.42$, $p = 0.004$, respectively).

For each stimulus face we calculated the mean estimated aggression across participants and correlated that with both the facial width-to-height ratio and actual aggression of the stimulus faces. The mean estimated aggression of each stimulus face across participants was positively correlated with the facial width-to-height ratio ($r_{22} = 0.67$, $p < 0.001$) and with actual aggression ($r_{22} = 0.37$, $p = 0.08$) in the 'blur' condition (see left panel of figure 4a) but was not correlated with either facial width-to-height ratio or actual aggression in the 'scramble' condition ($r_{22} = -0.10$, $p = 0.65$; and $r_{22} = -0.26$, $p = 0.21$, respectively—see left panel of figure 4b).

Most participants (87.5%) said that the eyes were the basis for their estimates of aggression. The mouth was said to be the basis by 44%, the eyebrows by 25%, and the shape of the face by 25%. No participant reported using any kind of configural information, including the facial width-to-height ratio.

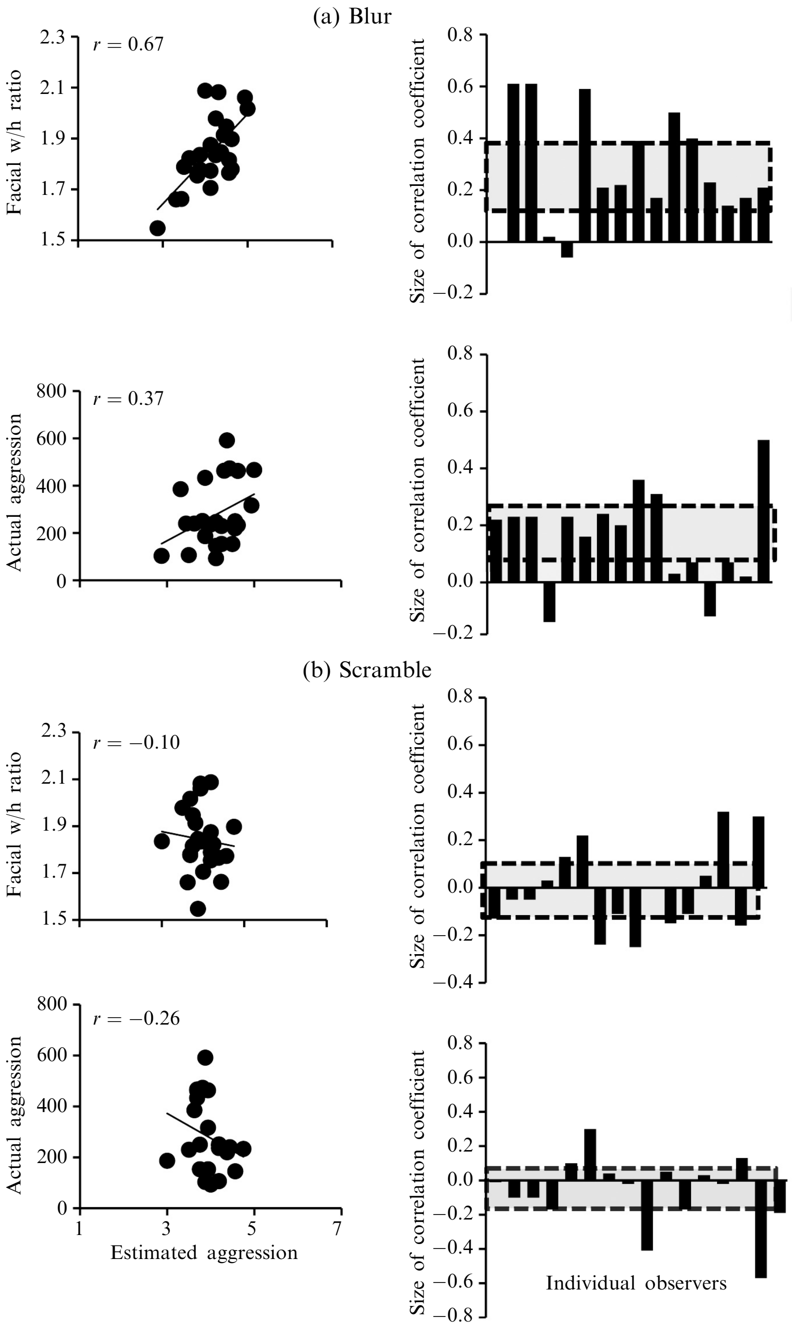

Figure 4. (a) Blur condition, left panel: relationship between mean estimated aggression across participants and facial width-to-height ratio (facial w/h ratio) (white background) and actual aggression (shaded background). Right panel: relationship for each individual observer between estimated aggression and facial width-to-height ratio (facial w/h ratio) (white background, mean r-value $= 0.28$, 95% CI $= 0.16-0.39$) and actual aggression (shaded background, mean r-value $= 0.16$, 95% CI $= 0.07-0.25$). (b) Scramble condition, left panel: relationship between mean estimated aggression across participants and facial width-to-height ratio (white background) and actual aggression (shaded background). Right panel: relationship for each individual observer between estimated aggression and facial width-to-height ratio (white background, mean r-value $= -0.01$, 95% CI $= -0.11-0.08$) and actual aggression (shaded background, mean r-value $= -0.07$, 95% CI $= -0.18-0.04$). Note. 39 ms exposure to each stimulus face.

2.4 *Experiment 1c*

Whereas participants in experiments 1a and 1b accurately judged aggression for 'blur' faces, the results for 'scramble' faces differed in the two experiments. Participants were accurate when 'scramble' faces were presented for 2000 ms and were preceded by exposures to unscrambled versions of the faces (experiment 1a), but they were inaccurate when the faces were presented for 39 ms and when the exposure to unscrambled faces was reduced (experiment 1b). In experiment 1c, we eliminated the pre-exposure to unmanipulated faces and increased presentation time to 2000 ms. If now the participants are not accurate in the 'scramble' condition, we can conclude that the accuracy in experiment 1a was the result of the pre-exposure to the unmanipulated faces and that their inaccuracy in experiment 1b was not the result of the short presentation time. A second group of participants was tested in the 'blur' condition as a direct comparison. We hypothesised that accuracy in the 'blur' condition would remain high in the absence of pre-exposure to the faces. Of course, this does not eliminate the possibility that the accuracy in the 'blur' condition in experiment 1b required some pre-exposure to the faces, although this is unlikely given how brief the pre-exposures were.

2.4.1 *Procedure.* We randomly assigned participants to either the 'blur' condition or to the 'scramble' condition (between-subject design). The participants viewed each face in either condition for 2000 ms. The order of faces was fully randomised within both conditions, and the presentation of a stimulus face was preceded by a central fixation cross that appeared for 500 ms. A face was then presented for 2000 ms, after which the question "How aggressive would this person be if provoked?" appeared on a black background along with a seven-point Likert scale (1—not at all aggressive, to 7—very aggressive). As in experiments 1a and 1b, participants were told how the aggressive behaviour of the men had been assessed before the presentation of stimulus faces.

2.4.2 *Results.* The estimates of aggression were consistent across individual observers in the 'blur' and in the 'scramble' conditions (Cronbach's $\alpha = 0.86$ and 0.71, respectively). For each participant we calculated the correlation between his/her estimate of aggression for the 24 faces, and both the facial width-to-height ratio and actual aggression of the stimulus faces. One-sample t-tests were computed to test the primary hypothesis that these correlations would be significantly different from the null hypothesis (ie no association) in the 'blur' but not the 'scramble' condition. The t-tests indicated that the estimates of aggression of individual participants were positively correlated with the facial width-to-height ratio and with actual aggression in the 'blur' condition ($t_9 = 4.94$, $p < 0.001$; and $t_9 = 6.18$, $p < 0.001$, respectively—see the right panel of figure 5a), but not in the 'scramble' condition ($t_9 = 2.05$, $p = 0.07$; and $t_9 = 1.05$, $p = 0.32$, respectively—see the right panel of figure 5b). For 'scramble' faces, the 95% confidence intervals for the relationship between estimated aggression and facial width-to-height ratio and estimated aggression and actual aggression included zero.

To compare the two experimental conditions, the correlations were converted to Fisher's z-scores, and then we computed independent sample t-tests on the Fisher's z transformed correlations. These t-tests indicated that the correlations between estimates of aggression and the facial width-to-height ratio, and estimates of aggression and actual aggression were significantly higher in the 'blur' condition than in the 'scramble' condition ($t_{18} = 3.23$, $p = 0.005$; and $t_{18} = 3.49$, $p = 0.003$, respectively).

For each stimulus face we calculated the mean estimated aggression across participants. The mean estimated aggression of each stimulus face was positively correlated with the facial width-to-height ratio and with actual aggression in the 'blur' condition ($r_{22} = 0.45$, $p = 0.03$; and $r_{22} = 0.49$, $p = 0.02$, respectively—see left panel of figure 5a), but not in the 'scramble' condition ($r_{22} = 0.15$, $p = 0.48$; and $r_{22} = 0.10$, $p = 0.64$, respectively—see left panel of figure 5b).

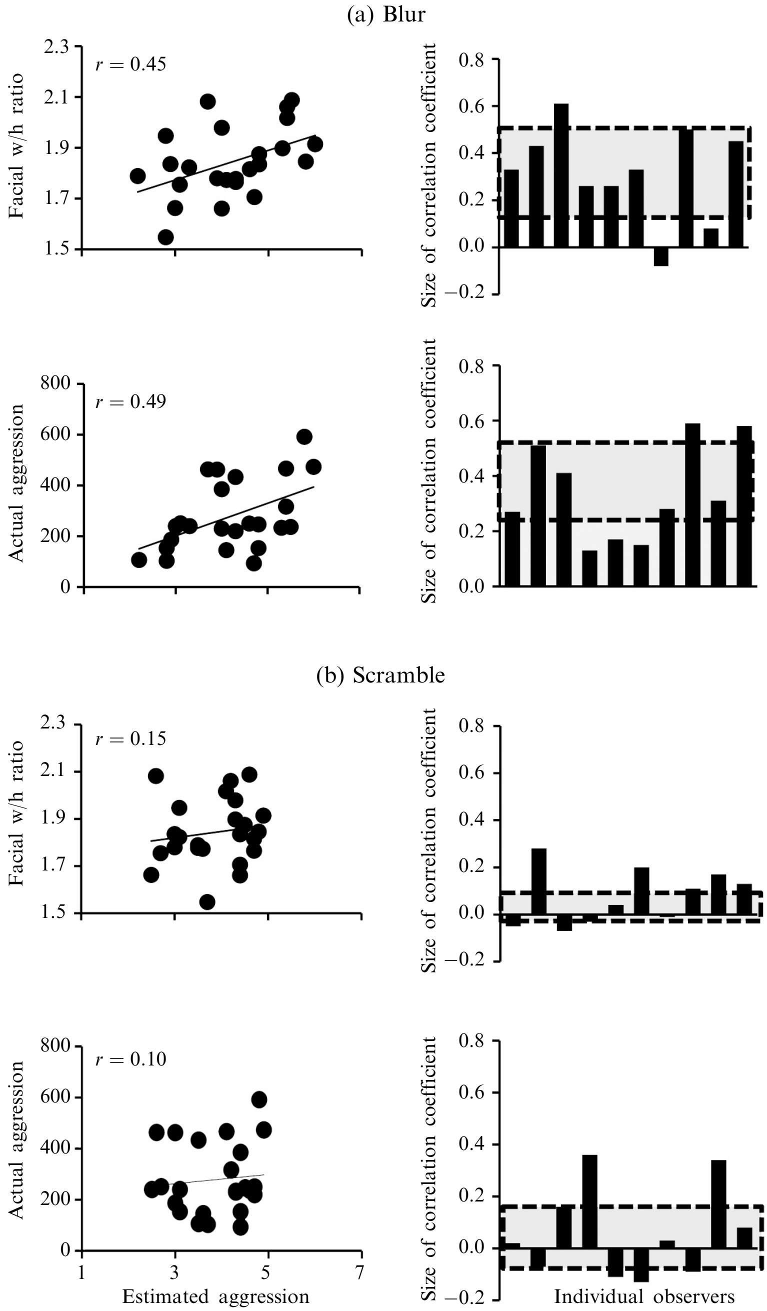

Figure 5. (a) Blur condition, left panel: relationship between mean estimated aggression across participants and facial width-to-height ratio (facial w/h ratio) (white background) and actual aggression (shaded background). Right panel: relationship for each individual observer between estimated aggression and facial width-to-height ratio (white background, mean r-value $= 0.32$, 95% CI $= 0.17-0.46$) and actual aggression (shaded background, mean r-value $= 0.34$, 95% CI $= 0.22-0.47$). (b) Scramble condition, left panel: relationship between mean estimated aggression across participants and facial width-to-height ratio (white background) and actual aggression (shaded background). Right panel: relationship for each individual observer between estimated aggression and facial width-to-height ratio (white background, mean r-value $= 0.08$, 95% CI $= -0.01-0.16$) and actual aggression (shaded background, mean r-value $= 0.06$, 95% CI $= -0.07$–0.19). Note. 2000 ms exposure to each stimulus face.

2.5 *Experiment 1d*

2.5.1 *Procedure.* The results of experiments 1b and 1c indicated that when individual features are presented in the absence of the facial width-to-height ratio ('scramble'), the accuracy in estimation of aggression drops to chance levels, even though most participants claimed to use individual features in the face as the basis for their estimates of aggression (results from experiment 1b). Thus, it appears that the use of the facial metric by the participants is implicit. In this experiment we investigated whether participants could explicitly judge variation in the facial width-to-height ratio.

Participants were taught the definition of the facial width-to-height ratio by pointing out the landmarks used for measurement on a photograph of a male face (not a stimulus face) for whom his facial width-to-height ratio was at the midpoint of the range of facial width-to-height ratios of the stimulus faces used in the current experiments. To ensure that participants understood the ratio, they were quizzed with questions such as whether the ratio would increase or decrease if the width of the face was held constant and the distance between brow and upper lip was increased. When the participants were able to answer accurately, the experimental stimuli were presented. The presentation of each unmanipulated stimulus face was preceded by a central fixation cross that appeared for 500 ms. A face was then presented for 39 ms, after which the question "How large is the width-to-height ratio?" appeared on a black background along with a seven-point Likert scale (1 – not at all large, to 7 – very large).

2.5.2 *Results.* Despite the restricted range of estimates (eg ratings between 3 and 5), estimates of the facial width-to-height ratio were positively correlated with the actual facial width-to-height ratio ($r = 0.70$, $p < 0.001$). A one-sample t-test was calculated to test the hypothesis that correlations between estimates of the facial width-to-height ratio of individual participants and the actual facial width-to-height ratio would be significantly different from the null hypothesis (ie no association). The t-test indicated that estimates of the facial width-to-height ratio of individual participants were positively correlated with the actual facial width-to-height ratio ($t_{11} = 3.96$, $p = 0.002$).

2.6 *General discussion—Experiments 1a – 1d*

The results from experiments 1a, 1b, and 1c provide several replications of our previous report that observers can make accurate estimates of aggression from emotionally neutral faces (Carré et al 2009). The results further suggest that the facial width-to-height ratio is a basis for the accuracy of judgments of aggression: estimates remained accurate when manipulations were made to a face in which the facial width-to-height ratio was maintained but external features such as jaw-line and forehead were eliminated (by cropping) or internal featural cues were reduced (by blurring). In contrast, estimates of aggression were disrupted when the facial width-to-height ratio was eliminated by scrambling the facial features. Although estimates did remain above chance in experiment 1a in the 'scramble' condition, they dropped to chance levels in experiments 1b and 1c. In experiment 1a, there was a 6 to 8 s exposure to the unscrambled faces (ie unmanipulated pre-exposure, crop, and blur conditions) which may have provided opportunity for learning to have occurred. When prior exposure to faces was reduced (experiment 1b) or removed (experiment 1c), accuracy in estimates of aggression from scrambled faces dropped to chance levels. Thus, overall, estimates of aggression are most accurate when all facial features (even blurred) are presented in their canonical arrangement, allowing for perception of the facial width-to-height ratio with, at most, a small contribution from the appearance of individual features. Although estimates of aggression were correlated with both facial width and facial height (components of the facial width-to-height ratio), these correlations were weaker than that between estimates of aggression and the facial width-to-height ratio (experiment 1a); thus the facial width-to-height ratio is both more diagnostic (Weston et al 2007; Carré et al 2009) and more relied upon by perceivers than the two individual metrics from which it is derived. It appears that the use of the facial width-to-height ratio by observers to form estimates of propensity for aggression was implicit, because the observers reported that they used features such as the eyes or mouth in making their judgments. Nonetheless, it is impressive that with just 39 ms exposure to each face (experiment 1d), observers can make highly accurate ($r = 0.70$) estimates of the facial width-to-height ratio.

The above experiments provide strong evidence that the facial width-to-height ratio is a critical factor guiding estimates of aggression. Importantly, when configural properties of the face were available (eg crop and blurred conditions), we consistently observed a relationship between estimates of aggression and actual aggression, as well as between estimates of aggression and facial width-to-height ratio. The consistency with which these relationships were observed confirms the robustness of the facial width-to-height ratio as a cue to a target's propensity for aggression. Nonetheless, it remains possible that other facial metrics, which may or may not be correlated with the facial width-to-height ratio, influence estimates of aggression. One or more of such metrics may be why there is some accuracy in estimates of aggression in the 'scrambled' condition, albeit only if previous opportunities to view the stimulus faces are provided (experiment 1a). In experiment 2, we used a data-driven statistical model of face representation (Blanz and Vetter 1999; Singular Inversions 2006) to create digitised versions of our original stimuli and to examine the extent to which other facial metrics are correlated with estimates of aggression and whether these metrics are correlated with the facial width-to-height ratio.

3 Experiment 2. Relationship between facial cues other than the facial width-to-height ratio and estimates of aggression

3.1 *Method*

3.1.1 *Procedure.* Here, we used a different approach to investigate the extent to which other facial metrics may explain variability in estimates of aggression. To do so, we used Facegen software, a data-driven statistical model of face representation that provides numerical values for 61 facial metrics (see table 1, and see Facegen.com for a more complete description of each facial metric—Blanz and Vetter 1999; Singular Inversions 2006). We uploaded each of our unmanipulated stimulus photographs used in experiment 1 into Facegen's PhotoFit option (see figure 6 for examples). For each of the 61 facial metrics, the software generates a z-score from Facegen's normal distribution of faces (based on laser scans of human faces). We also measured the facial width-to-height ratio from each digitised image generated by Facegen from our uploaded facial stimuli (using NIH ImageJ software) to ensure that the digitised faces maintained the same facial width-to-height ratio as the photographs used for our stimuli. We then computed zero-order correlation coefficients to examine the extent to which the facial metrics generated by Facegen correlated with the estimated aggression and with the facial width-to-height ratio. We also computed partial correlations to examine the extent to which any of these facial metrics would explain additional variance of estimated aggression over and above that explained when using only the facial width-to-height ratio. An ideal approach would be to conduct multivariate analyses to examine the relationship among the variables as predictors of estimates of aggression. But this possibility was limited by the small sample of faces for which aggression had been measured as well as by the high correlations among predictor variables (ie multicollinearity). Therefore, we took a conservative approach by not adjusting our α level for multiple correlations, thereby increasing the opportunity for other facial metrics to explain variance in estimated aggression. We then included those variables for which there was a significant partial correlation and the facial width-to-height ratio in a multiple regression analysis to predict estimates of aggression.

Table 1. Relationship between the facial width-to-height ratio (face w/h ratio), the 61 Facegen facial metrics and observer estimates of aggression. Zero-order (*r*) and partial (p*r*) correlations are reported for the relationship between estimated aggression and the facial metrics. Significant correlations are in bold.

Facegen metrics	Estimated aggression		Face w/h ratio	Facegen metrics	Estimated aggression		Face w/h ratio
	r	p*r*	*r*		*r*	p*r*	*r*
Brow ridge (high/low)	**0.47**	0.20	**0.59**	Jaw (wide/thin)	−0.26	0.03	**−0.48**
Brow ridge inner (down/up)	**−0.64**	−0.41	**−0.70**	Jaw neck slope (high/low)	0.29	0.16	0.29
Brow ridge outer (up/down)	0.32	0.08	**0.45**	Jawline (concave/convex)	0.22	0.07	0.29
Cheekbone (low/high)	0.11	−0.08	0.26	Lips (deflated/inflated)	−0.07	−0.14	0.08
Cheekbone (shallow/pronounced)	0.09	−0.09	0.28	Lips (large/small)	0.02	0.04	−0.01
Cheekbone (thin/wide)	0.25	−0.11	**0.56**	Lips (puckered/retracted)	−0.29	−0.23	−0.18
Cheeks (concave/convex)	0.16	−0.05	0.33	Lips (thin/thick)	0.02	0.13	−0.16
Cheeks (round/gaunt)	−0.06	0.07	−0.20	Mouth (drawn/pursed)	0.16	0.25	−0.08
Chin (forward/backward)	−0.17	0.23	**−0.55**	Mouth (happy/sad)	0.02	0.19	−0.21
Chin (pronounced/recessed)	−0.01	0.13	−0.20	Mouth (protruding/retracted)	−0.01	0.02	−0.04
Chin (retracted/jutting)	0.18	−0.05	0.37	Mouth tilt (up/down)	−0.23	−0.23	−0.06
Chin (shallow/deep)	**0.47**	0.16	**0.63**	Mouth (underbite/overbite)	−0.12	−0.14	−0.01
Chin (small/large)	−0.17	−0.11	−0.14	Mouth (up/down)	−0.21	−0.10	−0.22
Chin (tall/short)	0.19	−0.03	0.37	Mouth (wide/thin)	0.18	0.19	0.05
Chin (wide/thin)	0.33	0.09	0.44	Mouth–chin (short/long)	0.09	−0.05	0.21
Eyes (down/up)	−0.28	−0.05	−0.41	Nose bridge (shallow/deep)	−0.11	−0.03	−0.16
Eyes (small/large)	−0.27	−0.27	−0.10	Nose bridge (short/long)	−0.13	−0.01	−0.21
Eyes tilt (inward/outward)	−0.22	0.004	−0.37	Nose (down/up)	0.01	−0.08	0.11
Eyes (apart/together)	0.16	−0.21	**0.52**	Nose (flat/pointed)	0.01	0.02	0.01
Face (brow-nose-chin ratio)	**0.59**	0.35	**0.64**	Nose nostril tilt (down/up)	−0.06	−0.19	0.16
Face (forehead-sellion-nose ratio)	0.16	0.01	0.27	Nose nostril (small/large)	0.05	−0.03	0.13
Face (heavy/light)	**−0.54**	−0.24	**−0.67**	Nose nostril (wide/thin)	0.12	**0.41**	0.34
Face (round/gaunt)	−0.35	−0.13	**−0.43**	Nose region (concave/convex)	−0.02	0.20	−0.30
Face (tall/short)	**0.48**	0.17	**0.65**	Nose sellion (down/up)	−0.06	−0.08	−0.21
Face (up/down)	0.32	0.17	0.33	Nose sellion (shallow/deep)	−0.34	−0.20	−0.33
Face (wide/thin)	**0.45**	0.20	**0.54**	Nose sellion2 (shallow/deep)	−0.31	−0.24	−0.21
Forehead (small/large)	−0.13	−0.28	0.16	Nose sellion (thin/wide)	−0.22	−0.25	−0.04
Forehead (tall/short)	0.21	0.21	0.07	Nose (short/long)	0.05	0.16	−0.13
Forehead tilt (forward/backward)	−0.26	**−0.41**	0.13	Nose tilt (down/up)	0.01	−0.09	0.14
Head (thin/wide)	0.25	−0.09	**0.53**	Temples (thin/wide)	**−0.53**	**−0.47**	0.28
Jaw (retracted/jutting)	0.18	−0.05	0.37				

3.1.2 *Results.* The facial width-to-height ratios calculated from the digitised stimulus faces were highly correlated with the facial width-to-height ratios calculated from the photographs of our original stimulus faces ($r = 0.97$, $p < 0.0001$). Furthermore, the facial width-to-height ratio from the digitised faces was positively correlated with observer estimates of aggression obtained in a previous experiment ($r = 0.58$, $p = 0.003$—see Carré et al 2009). 8 of the 61 correlations between Facegen-calculated facial metrics and estimates of aggression were significant, 7 of which were also highly correlated with the facial width-to-height ratio (see table 1 for correlation coefficients); most of these facial metrics are components of the facial width-to-height ratio (eg brow ridge high/low, face tall/short, face wide/thin). Several additional facial metrics were correlated with the facial width-to-height ratio, but not estimates of aggression (eg outer brow ridge up/down, chin foreward/backward, jaw wide/thin). After controlling for the variance in estimates of aggression accounted for by the facial width-to-height ratio, only 3 facial metrics remained significant at the $p < 0.05$ level (forehead tilt, nostril width, and temple width). The multiple regression analysis that included these three metrics and the facial width-to-height ratio accounted for 60% of the variance

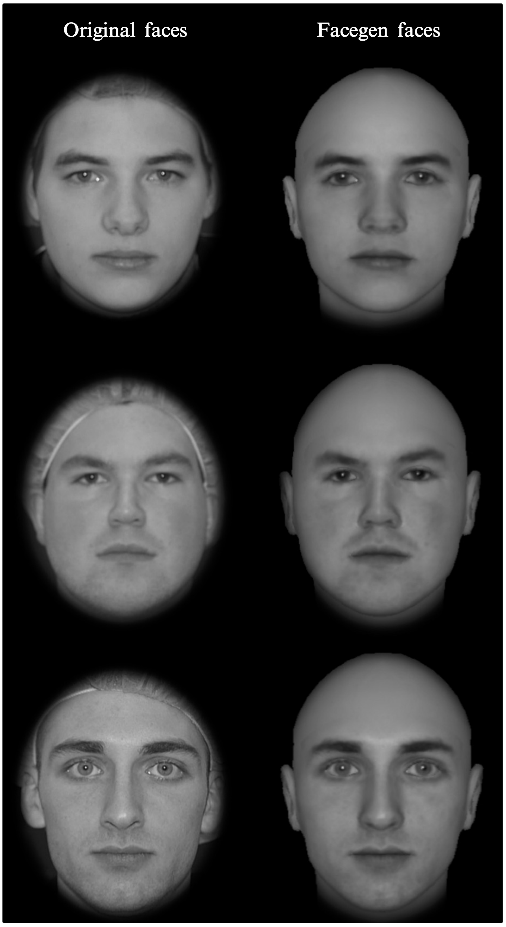

Figure 6. Examples of Facegen faces created using the original stimulus faces.

in estimates of aggression ($R^2 = 0.60$, $F_{4,19} = 7.23$, $p < 0.001$), with facial width-to-height ratio the only metric that made a significant unique contribution to the prediction ($p = 0.002$).

3.2 *Discussion*

The results obtained with Facegen indicate that other facial metrics are correlated with estimates of aggression. Interestingly, 7 of the 8 metrics that had zero-order correlations with estimates of aggression were also correlated with the facial width-to-height ratio. Nonetheless, after accounting for the variance in estimates of aggression explained by the facial width-to-height ratio, only 3 facial metrics emerged as significant predictors of estimates of aggression. Moreover, when these variables were simultaneously entered into a linear regression analysis, along with the facial width-to-height ratio, it was only the facial width-to-height ratio that made a significant unique contribution to the prediction of estimated aggression.

In the next experiment, we investigated whether the results of a relationship between facial width-to-height ratio and estimates of aggression that we obtained with our facial stimuli would extend to a novel set of stimulus faces. These computer-generated faces were made to vary according to the extent to which they were perceived as threatening (threat dimension), a variable strongly associated with aggression. A wealth of research indicates that people evaluate others along two primary dimensions; warmth and competence (see Fiske et al 2007, for review). The warmth dimension is composed of traits that are related to the perceived intent of the target, including traits such as friendly, trustworthy, and threatening. From an evolutionary perspective, warmth judgments are primary, as they enable observers to gauge another's intentions (positive or negative), which, through its influence on approach/avoidance behaviour, may have important consequences for survival (Fiske et al 2007). Judgments of aggression (as in Carré et al 2009, and experiments 1a – 1c) and threat (as in Oosterhof and Todorov 2008) would emerge as traits loading on the negative pole of the warmth dimension. Thus, a relationship between faces that vary in threat and the facial width-to-height ratio would provide a conceptual replication of our previous findings of a relationship between estimates of aggression and the facial width-to-height ratio.

4 Experiment 3. Relationship between the facial width-to-height ratio and a computer model of facial threat

4.1 *Method*

4.1.1 *Stimuli.* To investigate whether our results were limited to our real stimulus faces, or whether they would generalise to a novel set of synthetic stimulus faces, we used a database of faces that were freely available on Alexander Todorov's Social Cognition and Social Neuroscience Laboratory website. These faces were created with Facegen software to vary according to the extent to which they are perceived as threatening (threat dimension), a variable closely associated with aggression. This allowed us to ask whether faces generated with Facegen software to vary along the dimension of threat differ accordingly in facial width-to-height ratio. This database consists of 25 Caucasian faces that were generated with Facegen (Facegen Modeller program version 3.1, http://facegen.com) and from which seven versions of each facial identity were made to vary on the dimension of threat (Oosterhof and Todorov 2008). This provides a database of 175 faces (7 versions of 25 facial identities varying from -3 SD = not at all threatening to $+3$ SD = very threatening). In brief, to build their model, Oosterhof and Todorov (2008) asked participants to judge a series of computer-generated 3-D faces on the degree to which they appeared threatening, and then used the mean trait judgments made by observers to find vectors in the 50-dimensional face space in which direction is optimal in changing 'threat'. Next, Oosterhof and Todorov (2008) generated a new set of emotionally neutral faces and manipulated the facial metrics that were found to be most strongly correlated with 'threat'. Validation studies indicated that observer ratings tracked the ratings predicted by their computer model (Oosterhof and Todorov 2008). The question we asked was whether faces that were considered most threatening had the largest facial width-to-height ratio.

4.1.2 *Procedure and statistics.* The 175 face images in the database were re-coded so that measurements could be made blind of the level of threat of an individual image. Facial width-to-height ratios were then calculated with NIH ImageJ software as described above (see experiment 1a). The relationship between the facial width-to-height ratio and score on the threat dimension for the face images was then calculated for each facial identity with Pearson correlations and with a repeated-measures ANOVA.

4.1.3 *Results.* A repeated-measures ANOVA was computed to examine the extent to which facial width-to-height ratio would differ as a function of the threat dimension of the faces. The facial width-to-height ratio tracked the level of threat from individual facial identities ($F_{\text{linear},1,24} = 120.99$, $p < 0.001$), although the significant quadratic trend showed that differences in facial width-to-height ratio as a function of threat level were more pronounced at the high end of the threat dimension ($F_{\text{quadratic},1,24} = 5.79$, $p = 0.02$). A one-sample t-test comparing correlations for individual faces to a null-value of zero showed that threat dimension was positively associated with the facial width-to-height ratio ($t_{24} = 19.54$, $p < 0.001$—see figures 7a and 7b). As shown in figure 7b, this effect was consistent across individual stimulus faces.

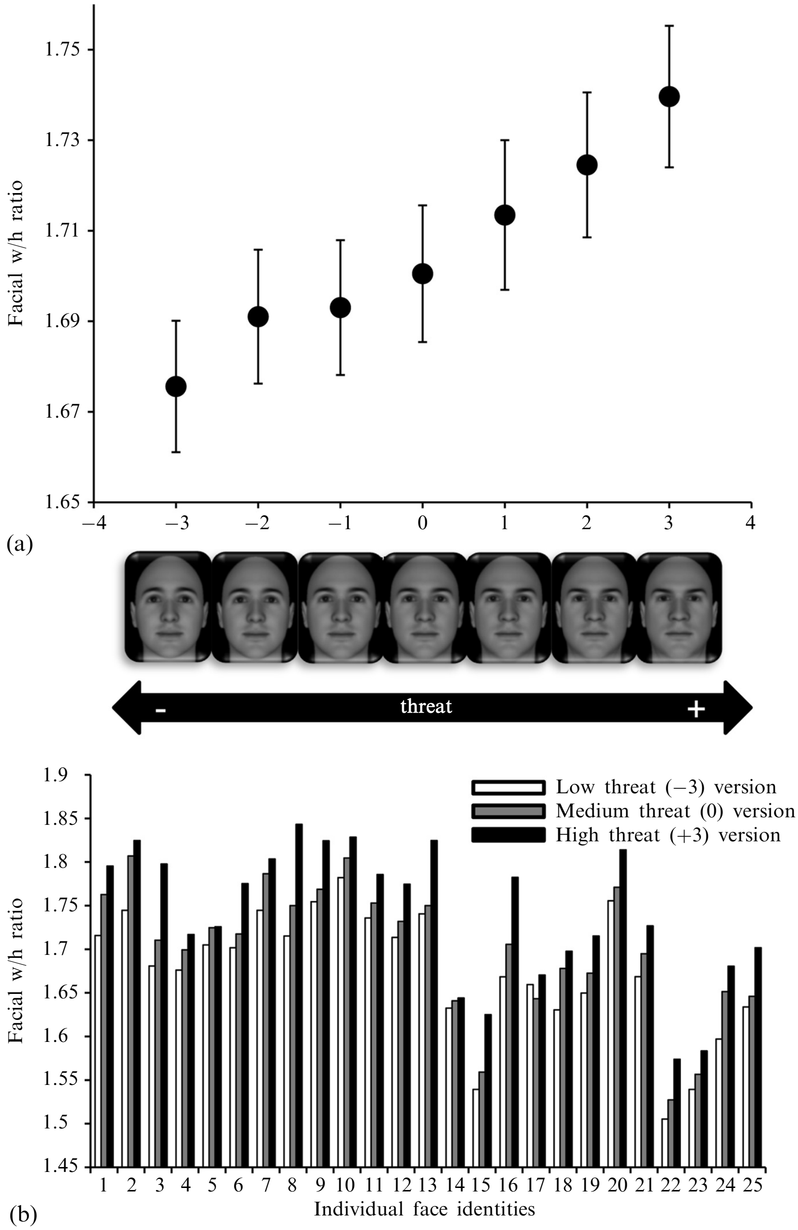

Figure 7. (a) Relationship between level of threat from emotionally neutral faces and facial width-to-height ratio (facial w/h ratio). (b) Relationship between level of threat from each individual facial identity ($n = 25$) and facial width-to-height ratio.

4.2 *Discussion*

The results of this experiment on a novel set of computer-generated stimulus faces add further support to the hypothesis that the facial width-to-height ratio is used by perceivers to predict aggressive potential. As each of the 25 face models varied along the dimension of threat, the facial width-to-height ratio also increased (see figure 7b). This finding indicates that the facial width-to-height ratio is associated with the degree of perceived threat, and also illustrates that the relationship between facial width-to-height ratio and personality extends beyond our original stimuli.

5 General discussion

In a previous study (Carré and McCormick 2008a) we reported that variation in the facial width-to-height ratio of male faces is diagnostic of propensity for aggression. The three sets of experiments reported here lend substantial support to the hypothesis that the facial width-to-height ratio is a cue used by perceivers to infer another's propensity for aggressive behaviour (Carré et al 2009). Perhaps the most impressive result from the current experiments is the consistency among observers within and across experiments in ratings of aggression in the stimulus faces. In addition, observer estimates of aggression remained accurate when faces were cropped and when featural properties of the face were made less salient by blurring (both of which maintain the integrity of the facial width-to-height ratio) and these estimates were highly correlated with the facial width-to-height ratio. The results obtained from the crop manipulations suggest that facial features above the eyebrows and below the upper lips are not required for accurate estimates of aggression. This is important because previous studies have suggested that variations in facial features such as jaw width, size of forehead, and size of lips are associated with perceptions of masculinity, dominance, and physical strength (see Keating et al 1981; McArthur and Apatow 1983/1984; Keating 1985; Perrett et al 1998). Thus, although ratings of propensity for aggression, masculinity, and dominance are conceptually related to each other, ratings of such traits may be influenced by different, separable facial cues. On the other hand, disrupting configural properties of the face (including the facial width-to-height ratio) significantly reduced accuracy in estimates of aggression, suggesting that individual features presented outside of the face context play a small (experiment 1a) or no (experiments 1b and 1c) role in estimates of aggression. Although individual facial characteristics may contribute to estimates of aggression in intact stimuli collectively, our results demonstrate that the facial width-to-height ratio is both diagnostic of aggressive potential (Carré and McCormick 2008a) and is used accurately by perceivers when estimating propensity for aggression (Carré et al 2009, and the present experiments 1a – 1c). This is particularly striking in light of the sexual dimorphism in the facial width-to-height ratio that emerges at puberty (Weston et al 2007).

Although research has demonstrated that both configural and featural properties of the face are important for face recognition (see Schwaninger et al 2002; Hayward et al 2008), it is possible that the mechanisms underlying face recognition and formation of trait impressions are not the same. In fact, Todorov and Duchaine (2008) recently reported that developmental prosopagnosics, that are seriously impaired in their ability to recognise individual face identities, are nonetheless capable of discriminating between trustworthy and untrustworthy-looking faces. Notably, ratings of trustworthiness are negatively correlated with ratings of aggressive potential ($r = -0.90$) (Carré et al 2009). Given the importance of configural properties of the face in generating accurate estimates of aggression, and the finding that at least one kind of configural face processing—sensitivity to differences among faces in the spacing of features—develops slowly in children (see Freire and Lee 2001; Mondloch et al 2002), it will be of interest to examine development of children's accuracy in estimating aggression from emotionally neutral faces.

The Facegen software calculations of 61 metrics in each stimulus face in experiment 2 further support the hypothesis that the facial width-to-height ratio is a basis of observers' accurate estimates of aggression, in that only those metrics that are a basis of the facial width-to-height ratio were associated with estimates of aggression. Although 8 other facial metrics were correlated with estimates of aggression, 7 of them were also highly correlated with the facial width-to-height ratio. Moreover, our crop manipulations reduced the pool of potentially required metrics for estimating aggression to 4, and these 4 (brow ridge high/low, inner brow ridge down/up, face tall/short, face wide/thin) all are involved in deriving the facial width-to-height ratio. Further, the finding of a relationship between facial width-to-height ratio and perceived threat in a database of faces generated in an independent lab (Oosterhof and Todorov 2008) highlights the generalisability of the findings (Oosterhof and Todorov 2008; Carré et al 2009). We found that the relationship between the facial width-to-height ratio and perceived threat was both linear and quadratic. More specifically, differences in the facial width-to-height ratio were more pronounced at the high end of the threat dimension compared to the low end of the dimension (see figure 7a). This finding parallels those obtained by Oosterhof and Todorov (2008) wherein observers' ratings of the degree to which faces appeared 'threatening' were more sensitive to change at the high-threat relative to low-threat end of the dimension.

The fact that observers are consistent in their estimates of threat (Oosterhof and Todorov 2008) and of aggression (the current experiments and Carré et al 2009) from emotionally neutral faces suggests that the perceptual system may have evolved to detect subtle static facial cues that may have been important in guiding adaptive social behaviours. Thus, one possible explanation for the current findings (and those obtained by Carré et al 2009) is that the facial width-to-height ratio may represent a more subtle signal of aggressive potential, and that a more overt signal, such as an angry facial expression, may serve to amplify this signal. Thus, the ability to make accurate estimates of aggression may be related, in part, to an overgeneralisation of emotional expressions (McArthur 1982; Montepare and Dobish 2003; Zebrowitz and Montepare 2008), whereby individuals that have facial metrics that resemble a particular emotional state will be perceived as actually showing that emotion. Consistent with this proposition is the recent finding that decreasing the distance between the brow and lip (which would also increase the facial width-to-height ratio) increases the perception of anger from emotionally neutral faces (Neth and Martinez 2009). Notably, an angry facial expression consists of lowering the brow and raising the upper lip (Ekman 2007), a facial movement that would also increase the facial width-to-height ratio, and thus increase the perception of aggressive intent.

If trait judgments from emotionally neutral faces reflect an overgeneralisation of emotional expressions, individuals with emotion recognition deficits may be impaired in their ability to accurately estimate another's propensity for aggression. Many studies have found that individuals with bilateral amygdala damage are impaired in the recognition of facial emotional expressions (although deficits after amygdala damage are not exclusive to emotional recognition; see review by Adolphs 2008). For example, Adolphs and colleagues (1998) reported that patients with bilateral amygdala damage were impaired in their ability to discriminate between trustworthy and untrustworthy faces compared to age-matched controls. Importantly, face recognition remains intact among these patients, providing additional support for the idea that face recognition and generation of trait impressions are dissociable. More recently, evidence has emerged to suggest that amygdala reactivity is negatively correlated with the degree of trustworthiness from natural photographs (Winston et al 2002; Engell et al 2007; Said et al 2009) and computer-generated images (Todorov et al 2008a). Together, these findings indicate that the amygdala is preferentially engaged during valence evaluation of emotionally neutral faces (Todorov et al 2008b; Blasi et al 2009). Given the robust negative correlation between estimates of aggression and trustworthiness (eg Oosterhof and Todorov 2008; Carré et al 2009), and the strong correlation between estimates of aggression and the facial width-to-height ratio, it is reasonable to propose that faces with high facial width-to-height ratio would generate preferential activation in the amygdala.

Nevertheless, it is important to note that the effects of the facial width-to-height ratio do not depend upon changing perceived emotion in the face. Facial stimuli from the Todorov database were judged by observers to be emotionally neutral (Oosterhof and

Todorov 2008). Nevertheless, the perceptual system may have evolved to be especially sensitive to slight variations in static facial structures that are associated with propensity for aggressive behaviour (eg Carré and McCormick 2008a). Similar physical signals are used in other species as a means to gauge one's relative social status within the hierarchy (Tibbetts and Dale 2004; Setchell et al 2008), which may ultimately serve to guide adaptive social behaviour (eg approach or avoidance behaviour—Senar and Camerino 1998; Tibbetts and Lindsay 2008). The consistency with which estimates of aggression were correlated with both actual aggression and facial width-to-height ratio provide strong support for the hypothesis that the facial width-to-height ratio is an honest signal of propensity for aggression.

Acknowledgments. We thank Mark Vida and Alexandra Hatry for technical assistance and Michael Ashton for comments on an earlier version of this paper. CMM holds a Natural Sciences and Engineering Research Council (NSERC) Discovery Grant and a Social Sciences and Humanities Research Council (SSHRC) Grant. CJM holds an NSERC Discovery Grant and JMC holds an NSERC Canada Graduate Scholarship.

References

Adolphs R, 2008 "Fear, faces and the human amygdala" *Current Opinion in Neurobiology* **18** 166 – 172

Adolphs R, Tranel D, Damasio A R, 1998 "The human amygdala in social judgment" *Nature* **393** 474

Ballew C C, Todorov A, 2007 "Predicting political elections from rapid and unreflective face judgments" *Proceedings of the National Academy of Sciences of the USA* **104** 17948 – 17953

Bar M, Neta M, Linz H, 2006 "Very first impressions" *Emotion* **6** 268 – 278

Baron R, Richardson D, 1994 *Human Aggression* second edition (New York: Plenum Press) page 7

Blair I V, Judd C M, Chapleau K M, 2004 "The influence of Afrocentric facial features in criminal sentencing" *Psychological Science* **15** 674 – 679

Blanz V, Vetter T, 1999 "A morphable model for the synthesis of 3D faces", in *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques* (Los Angeles: Addison Wesley Longman) pp 187 – 194

Blasi G, Hariri A R, Alce G, Taurisano P, Sambataro F, Das S, Bertoloni A, Weinberger D R, Mattay V S, 2009 "Preferential amygdala reactivity to the negative assessment of neutral faces" *Biological Psychiatry* **66** 847 – 853

Boothroyd L G, Jones B C, Burt D M, DeBruine L M, Perrett D I, 2008 "Facial correlates of sociosexuality" *Evolution and Human Behavior* **29** 211 – 218

Carré J M, McCormick C M, 2008a "In your face: Facial metrics predict aggressive behaviour in the laboratory and in varsity and professional hockey players" *Proceedings of the Royal Society of London, Section B* **275** 2651 – 2656

Carré J M, McCormick C M, 2008b "Aggressive behavior and change in salivary testosterone concentrations predict willingness to engage in a competitive task" *Hormones & Behavior* **54** 403 – 409

Carré J M, McCormick C M, Mondloch C J, 2009 "Facial structure is a reliable cue of aggressive behavior" *Psychological Science* **20** 1194 – 1198

Cherek D, Schnapp W, Moeller F, Dougherty D, 1996 "Laboratory measures of aggressive responding in male parolees with violent and non-violent histories" *Aggressive Behavior* **22** 27 – 36

Christiansen K, Winkler E M, 1992 "Hormonal, anthropometrical, and behavioral correlates of physical aggression in Kung San men of Namibia" *Aggressive Behavior* **18** 271 – 280

Collishaw S M, Hole G J, 2000 "Featural and configurational processes in the recognition of faces of different familiarity" *Perception* **29** 893 – 909

Ekman P, 2007 *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life* second edition (New York: Henry Holt)

Engell A D, Haxby J V, Todorov A, 2007 "Implicit trustworthiness decisions: automatic coding of face properties in the human amygdala" *Journal of Cognitive Neuroscience* **19** 1508 – 1519

Fiske S T, Cuddy A M, Glick P, 2007 "Universal dimensions of social cognition: warmth and competence" *Trends in Cognitive Sciences* **11** 77 – 83

Freire A, Lee K, 2001 "Face recognition in 4 to 7 year-olds: Processing of configural, featural, and paraphernalia information" *Journal of Experimental Child Psychology* **80** 347 – 371

Hassin R, Trope Y, 2000 "Facing faces: Studies on the cognitive aspects of physiognomy" *Journal of Personality and Social Psychology* **78** 1174 – 1194

Hayward W G, Rhodes G, Schwaninger A, 2008 "An own-race advantage for components as well as configurations in face recognition" *Cognition* **106** 1017 – 1027

Keating C F, 1985 "Gender and the physiognomy of dominance and attractiveness" *Social Psychology Quarterly* **48** 61 – 70

Keating C F, Mazur A, Segall M, 1981 "A cross-cultural exploration of physiognomic traits of dominance and happiness" *Ethology and Sociobiology* **2** 41 – 48

Little A C, Burris R P, Jones B C, Roberts S C, 2007 "Facial appearance affects voting decisions" *Evolution and Human Behavior* **28** 18 – 27

Little A C, Perrett D I, 2007 "Using composite images to assess accuracy in personality attribution to faces" *British Journal of Psychology* **98** 111 – 126

McArthur L, 1982 "Judging a book by its cover: A cognitive analysis of the relationship between physical appearance and stereotyping", in *Cognitive Social Psychology* Eds A Hastorf, A Isen (New York: Elsevier/North-Holland) pp 149 – 211

McArthur L Z, Apatow K, 1983/1984 "Impressions of baby-faced adults" *Social Cognition* **2** 315 – 342

Mazur A, Mazur J, Keating C F, 1984 "Military rank attainment of a West Point class: Effects of cadets' physical features" *American Journal of Sociology* **90** 125 – 150

Mondloch C J, Le Grand R, Maurer D, 2002 "Configural face processing develops more slowly than featural face processing" *Perception* **31** 553 – 566

Montepare J M, Dobish H, 2003 "The contribution of emotion perception and their overgeneralisations to trait impressions" *Journal of Nonverbal Behavior* **27** 237 – 254

Neth N, Martinez A M, 2009 "Emotion perception in emotionless face images suggests a norm-based representation" *Journal of Vision* **9** 1 – 11

Oosterhof N N, Todorov A, 2008 "The functional basis of face evaluation" *Proceedings of the National Academy of Sciences of the USA* **105** 11087 – 11092

Penton-Voak I S, Pound N, Little A C, Perrett D I, 2006 "Personality judgments from natural and composite facial images: more evidence for a 'kernel of truth' in social perception" *Social Cognition* **24** 490 – 524

Perrett D I, Lee K J, Penton-Voak I, Rowland D, Yoshikawa S, Burt D M, Henzi S P, Castles D L, Akamatsu S, 1998 "Effects of sexual dimorphism on facial attractiveness" *Nature* **394** 884 – 887

Roney J R, Hanson K N, Durante K M, Maestripieri D, 2006 "Reading men's faces: women's mate attractiveness judgments track men's testosterone and interest in infant" *Proceedings of the Royal Society of London, Section B* **273** 2169 – 2175

Rule N O, Ambady N, 2008a "Brief exposures: Male sexual orientation is accurately perceived at 50 ms" *Journal of Experimental Social Psychology* **44** 1100 – 1105

Rule N O, Ambady N, 2008b "The face of success: Inferences from chief executive officers' appearance predict company profits" *Psychological Science* **19** 109 – 111

Rule N O, Ambady N, Adams R B Jr, Macrae C N, 2008 "Accuracy and awareness in the perception and categorization of male sexual orientation" *Journal of Personality and Social Psychology* **95** 1019 – 1028

Said C P, Baron S G, Todorov A, 2009 "Nonlinear amygdala response to face trustworthiness: contributions of high and low spatial frequency information" *Journal of Cognitive Neuroscience* **21** 519 – 528

Schwaninger A, Lobmeier J S, Collishaw S M, 2002 "Role of featural and configural information in familiar and unfamiliar face recognition" *Lecture Notes in Computer Science* **2525** 643 – 650

Sell A, Cosmides L, Tooby J, Sznycer D, Rueden C von, Gurven M, 2009 "Human adaptations for the visual assessment of strength and fighting ability from the body and face" *Proceedings of the Royal Society of London, Section B* **276** 575 – 584

Senar J C, Camerino M, 1998 "Status signalling and the ability to recognize dominants: an experiment with siskins (*Carduelis spinus*)" *Proceedings of the Royal Society of London, Section B* **265** 1515 – 1520

Setchell J M, Smith T, Wickings E J, Knapp L A, 2008 "Social correlates of testosterone and ornamentation in male mandrills" *Hormones & Behavior* **54** 365 – 372

Singular Inversions, 2006 Facegen Software (Version 3.3.1), http://www.facegen.com, retrieved on 20 March 2009

Stillman T F, Maner J K, 2009 "A sharp eye for her SOI: perception and misperception of female sociosexuality at zero acquaintance" *Evolution and Human Behavior* **30** 124 – 130

Tibbetts E A, Dale J, 2004 "A socially enforced signal of quality in paper wasps" *Nature* **432** 218 – 222

Tibbetts E A, Lindsay R, 2008 "Visual signals of status and rival assessment in *Polistes dominulus* paper wasps" *Biology Letters* **4** 237 – 239

Todorov A, Baron S G, Oosterhof N N, 2008a "Evaluating face trustworthiness: a model based approach" *Social, Cognitive and Affective Neuroscience* **3** 119 – 127
Todorov A, Duchaine B, 2008 "Reading trustworthiness in faces without recognizing faces" *Cognitive Neuropsychology* **25** 395 – 410
Todorov A, Mandisodza A N, Goren A, Hall C C, 2005 "Inferences of competence from faces predict election outcome" *Science* **308** 1623 – 1626
Todorov A, Said C P, Engell A D, Oosterhof N H, 2008b "Understanding evaluation of faces on social dimensions" *Trends in Cognitive Sciences* **12** 455 – 460
Weston E M, Friday A E, Lio P, 2007 "Biometric evidence that sexual selection has shaped the hominin face" *PLoS ONE* **2** 1 – 8
Willis J, Todorov A, 2006 "First impressions: Making up your mind after 100 ms exposure to a face" *Psychological Science* **17** 592 – 598
Winston J S, Strange B A, O'Doherty J, Dolan R J, 2002 "Automatic and intentional brain responses during evaluation of trustworthiness of faces" *Nature Neuroscience* **5** 277 – 283
Zebrowitz L A, 1997 *Reading Faces: Window to the Soul?* (Boulder, CO: Westview Press)
Zebrowitz L A, Montepare J M, 2008 "Social psychological face perception: Why appearance matters" *Social and Personality Psychology Compass* **2** 1497 – 1517

ISSN 0301-0066 (print) ISSN 1468-4233 (electronic)

PERCEPTION
VOLUME 39 2010

www.perceptionweb.com