THE FACIAL WIDTH-TO-HEIGHT RATIO AND ITS ROLE IN ADVERTISEMENTS AND ASSESSMENTS OF THREAT POTENTIAL

Shawn N. Geniole

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Department of Psychology
BROCK UNIVERSITY
St. Catharines, Ontario

Abstract

As do many species, humans visually assess the ability and propensity of others to cause trouble or harm (threat potential), although the mechanisms that guide this ability are unknown. One potential mechanism that may underlie advertisements and assessments of threat is the facial width-to-height ratio (face ratio). The overarching goal of this thesis was to test both the ecological validity of the face ratio (i.e., the extent to which it maps onto an individual's actual threat potential), and its utility in influencing observers' first impressions of traits related to threat potential. In Chapter 2, I found that men (*n* = 146) but not women (*n* = 76) with larger face ratios were more likely to cheat in a lottery for a cash prize than were men with smaller face ratios. In Chapter 3, to better identify the precise social function of the metric, I examined its differential association with two types of threat-related judgements, untrustworthiness and aggressiveness. The face ratio (*n* of faces = 141) was more strongly linked to observers' (*n* = 129) judgements of aggression than to their judgements of trust, although it is possible that this metric advertises threat potential more generally, of which aggression is a best indicator. In Chapter 4 (which extended some preliminary, additional findings from Chapter 3), I found that observers' (*n* = 56) judgements of aggression were strongly correlated with the face ratio (*n* of faces = 25) even when men were bearded, suggesting that this metric could have been operational in our ancestral past when interactions likely involved bearded men. In Chapter 5, I combined effect sizes from experiments conducted from several independent labs and identified significant (albeit weak) associations between the face ratio and actual threat behaviour, and significant (and stronger) associations between the face ratio and judgements of threat potential. Together, this body of work provides initial evidence that the face ratio, and sensitivity to it, may be part of an evolved system designed for advertising and assessing threat in humans, akin to threat assessment systems identified in other species.

Keywords: threat assessment; badge of status; signal; antisocial behaviour; aggression

Acknowledgements

First and foremost, I would like to thank my advisor, Dr. McCormick, for all of her support and guidance throughout my PhD. She made me strive to be as productive and as successful as possible in all aspects of graduate school, and I can't imagine achieving as much as I have without her. I would also like to thank my wife, Valentina, for all of her love and support, and our son Leo, for reminding us how important life is outside of academia. Other members of the McCormick lab (Matt, Jonathan, Travis, Marina, Elliott, and all of the amazing research assistants) and my good friends (and mentors) Chloe and Paul have also been incredibly helpful and supportive; thank you for everything. My committee members Dr. Busseri and Dr. Hafer have also provided many suggestions and comments that have improved the quality of this thesis. I can't thank you enough for your time and your support. Also, thank you Dr. Busseri for the additional mentorship and many important discussions throughout graduate school. Finally, I would like to thank my dad, my mom, and my second family, the Groves, for everything else along the way. I am so lucky to have such a loving and caring family.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1: General Introduction

Advertising and Assessing Threat Potential

Evidence from non-human animals. The ability for an organism to detect and assess threat in an environment is of great importance for survival and reproduction (Bar, Neta, & Linz, 2006; Blanchard, Griebel, Pobbe, & Blanchard, 2011). It is also beneficial for organisms to advertise their own threat potential[1]; doing so reduces the need for contest escalation and injury (Maynard Smith & Harper, 2003). For example, organisms favoured to win contests could advertise rather than behaviourally express their threat potential, thus saving energy, and organisms likely to lose contests could assess the conspecific's threat potential and submit/flee before sustaining possible injuries related to competition. Across many taxa, visual exposure to rival conspecifics before agonistic contests reduces the duration and the lethality of the contests [e.g., cichlids (*Cichlidae*); green swordtails (*Xiphophorus hellerii*); rainbow trout (*Oncorhynchus mykiss*); pigs (*Sus scrofa*); hamsters (*Mesocricetus brandti*), reviewed in Arnott & Elwood, 2009], providing support for the existence (and benefits) of such advertisement and assessment systems. Because of these advantages, selection pressures likely favoured mechanisms that allow for advertisements and assessments of threat potential.

What are some of the mechanisms through which non-human animals advertise and assess threat potential? In some species, researchers have identified physical markers or coloration patterns that correspond to the animal's dominance and aggressiveness. For

[1] According to the Merriam-Webster dictionary (http://www.merriam-webster.com/dictionary/threat), the term "threat" refers to "someone or something that could cause trouble, harm, etc". The term "threat potential" is used here to capture both an organism's ability and propensity to cause trouble or harm to other organisms. Therefore, the term can involve both physical qualities such as strength, skills such as fighting ability, and traits or tendencies, such as aggressiveness, untrustworthiness, prejudice, related to one's ability and propensity to cause trouble or harm. I use the terms "threat" and "threat potential" interchangeably throughout this thesis.

example, in male eland antelope (*Tragelaphus oryx*), the facemask becomes darker as the social dominance of the animal increases (Bro-Jørgensen & Beeston, 2015). In male mandrills (*Mandrillus sphinx*), facial coloration changes with dominance rank such that mandrills higher in dominance have redder faces than those lower in dominance (Setchell, Smith, Wickings, & Knapp, 2008). These findings highlight potential mechanisms through which these animals may advertise their threat potential. Nevertheless, for a mechanism to be involved in both the advertisement and assessment of threat potential, it should not only correspond with the threat potential of the animal, but also be perceivable by other conspecifics and modulate social behaviour.

Researchers have examined how cues of threat potential guide social behaviour in some species. For example, paper wasps (*Polistes dominulus*) avoid food sources guarded by wasps with more "broken" black facial patterns, which are indicative of greater threat potential (Tibbetts & Dale, 2004), and instead choose food sources guarded by wasps with less broken facial patterns (Tibbetts & Lindsay, 2008); great tits (*Parus major*) submissively avoid conspecifics with wider breast stripes, which are indicative of greater threat potential, but aggressively approach those with narrower ones (Järvi & Bakken, 1984), and; pipefish (*Syngnathus typhle*), when deciding whether to compete for a mate, are more deterred when the competitor has black cross stripes, which are indicative of greater threat potential, than when the competitor does not (Berglund & Rosenqvist, 2009). In all of these studies, the advertisements of threat potential were manipulated by the researchers (e.g., using pens and markers), providing experimental evidence that these cues function to alter the social behaviour of conspecifics.

Evidence from humans. Such threat assessment and advertising mechanisms would also be of great importance to humans. For example, accurately judging the threat potential of another individual would allow one to make appropriate decisions regarding approach/avoidance in social interaction or, if interactions become competitive or agonistic, appropriate decisions to fight or flee (Sell et al., 2009; Zebrowitz & Collins, 1997). As pointed out by Sell and colleagues (2009), the prevalence of aggression and violence throughout human history (e.g., Walker, 2001) likely had large selection pressures on humans, favouring perceptual and psychological mechanisms that advertise and assess threat potential. Although interpersonal violence was much higher in previous centuries (e.g., Eisner, 2003), many people still lose their lives to violence each day. In the year 2000 alone, for example, the World Health Organization reported that approximately 520,000 people lost their lives because of interpersonal violence, and another 310,000 people lost their lives from warfare (World Health Organization, 2002). Even when an altercation does not lead to the death of one of the parties, aggression and violence are still costly to both the aggressor and the victim (e.g., risk of injuries/death, legal ramifications)[2]. Therefore, selection should favour mechanisms in humans, as in other species, that allow for the advertisement and assessment of threat potential.

Much evidence for threat assessment in humans comes from studies in which participants are given information about a stranger’s voice or appearance and are asked to guess the threat potential of the individual. For example, certain components of human

---

[2] Threat behaviours also serve some beneficial social functions (e.g., helping an individual climb a status hierarchy, appropriate resources, or protect themselves against ongoing and future attacks, Daly & Wilson, 1988; Hawley & Vaughn, 2003), but to the extent that advertisements of threat potential are effective, they will often represent a more cost-efficient means of regulating social behaviour than acts of aggression and violence.

voice are related to indices of threat potential (e.g., upper body strength, aggression, Puts, Apicella, & Cárdenas, 2012) and, after listening to audio clips of a stranger's voice, participants form accurate estimates about the stranger's physical strength, fighting ability (e.g., Sell et al., 2010) and power (Berry, 1991) (for review, see Zebrowitz & Collins, 1997). Although accurately assessing threat based on the voice of other individuals would be beneficial for regulating social decisions, this information is not always available upon first encounters. Appearance cues, on the other hand, may be more readily available and also allow for more distant and thus earlier threat assessments, suggesting that humans may have developed cognitive mechanisms designed for extracting visual information about threat potential.

If advertising and assessing threat is important for reducing the costs of interpersonal conflict, then physical advertisements should be conspicuous[3] and visual systems should be sensitive to the location of these advertisements. One location of the body to which the human visual system is highly sensitive is the face. When scanning visual scenes, participants first locate the individual in the scene and then focus on the face, where they spend the majority of their looking time (compared to other parts of the body, e.g., Fletcher-Watson, Findlay, Leekam, & Benson, 2008), a visual preference that develops early, in infancy (e.g., Kwon, Setoodehnia, Baek, Luck, & Oakes, 2016). The face has been referred to as an organ of communication (Bruce & Young, 2012), and humans can rapidly extract information about sex, age, race, and emotions from the faces

[3] Other researchers have suggested that economy of effort is a central feature of advertisement and assessment systems, which some have interpreted to be in direct conflict with the idea that advertisements should be conspicuous or salient (reviewed in Scott-Phillips, 2008). Although more conspicuous displays may require greater energy to produce initially, they would likely reduce subsequent effort required for advertising threat potential and effort required for observers to perceive the advertisement, which may lead to net reductions in cost, over time, and represent a more efficient system, overall.

of strangers (reviewed in Bruce & Young, 2012; McGugin & Gauthier, 2013), thus highlighting the potential role of the face in advertisements of various traits.

Facial expressions may provide one channel through which humans communicate their behavioural intentions (Fridlund, 1994; McArthur & Baron, 1983; Yik & Russell, 1999), but these emotional expressions can be misrecognized (reviewed in Nelson & Russell, 2013; see also Aviezer, Trope, & Todorov, 2012), misleading (reviewed in Ekman, 2003), and suppressed to hide (e.g., having a poker-face) or enhanced to exaggerate intentions, traits, and abilities (e.g., Marsh, Adams, & Kleck, 2005; Marsh, Cardinale, Chentsova-Dutton, Grossman, & Krumpos, 2014; Sell, Cosmides, & Tooby, 2014). Not surprisingly, judgements based on dynamic cues in the face are less consistent than are those based on stable, static cues (Hehman, Flake, & Freeman, 2015). Therefore, humans may have developed the ability to detect and extract threat potential from static or structural (e.g., bone shape) cues in addition to dynamic ones (e.g., emotional expressions) in the face.

To investigate this possibility, researchers ask participants to make guesses about the threat potential of strangers based on photographs of the strangers' faces. Using such paradigms, researchers have provided some evidence for accurate assessments of threat potential. For example, perceptions of men's power (operationalized as a combination of perceived strength, assertiveness, invulnerability, and dominance) were associated with the men's actual self-reported assertiveness, social potency, aggression, and power (Berry, 1991; see also Berry, 1990); men high or low in Machiavellianism (which includes components of dominance and boldness) were identified as such from facial photographs, with identification rates higher than chance probability (Cherulnik, Way, Ames, & Hutto, 1981; see also Holtzman, 2011); when viewing photographs of male faces that were separated into high and low testosterone groups (testosterone is a hormone linked to dominance and aggression, for reviews see Archer, 2006; Carré, McCormick, & Hariri, 2011; Carré & Olmstead, 2015; Mazur & Booth, 1998), participants judged the high testosterone group of faces as looking more strong and dominant and as less friendly and good than the low testosterone group (Dabbs, 1997).

These examples of accuracy in threat assessment also extend to faces that are in the high end of the distribution regarding threat potential: Male fighters in the Ultimate Fighting Championship who were judged as looking more aggressive had higher win percentages than male fighters who were judged as looking less aggressive (Trebicky, Havlícek, Roberts, Little, & Kleisner, 2013), military cadets who were judged as looking more dominant were more likely to obtain higher military ranks later in their careers than were military cadets who were judged as looking less dominant (Mueller & Mazur, 1996), and violent sexual offenders were judged as looking more prone to violence than were non-violent sexual offenders (Stillman, Maner, & Baumeister, 2010). Although these later two studies also involved some faces that were not posed in neutral expressions, the effects did not appear to depend on the emotionality portrayed by the expression (in some cases, emotions led to inaccuracy rather than accuracy in these judgements, Stillman et al., 2010).

Although most of these findings involve samples of faces from Western, educated, industrialized, rich, and democratic societies (“WEIRD” societies, Henrich, Heine, & Norenzayan, 2010), they appear generalizable to other non-WEIRD populations as well. For example, male forager-farmers from the Bolivian Amazon (Tsimane’) who were judged as looking more dominant had greater bicep circumference, an index of fighting ability and aggressiveness (Gallup, O'Brien, White, & Wilson, 2010; Gallup, White, & Gallup, 2007; Muñoz-Reyes, Gil-Burmann, Fink, & Turiegano, 2012; Sell et al., 2009), than did those judged as looking less dominant (Undurraga et al., 2010). This association between perceptions of dominance and actual strength parallels that found involving faces and observers from WEIRD populations (Fink, Neave, & Seydel, 2007; see also Toscano, Schubert, & Sell, 2014; Windhager, Schaefer, & Fink, 2011). In another study, accurate estimates of strength were made based on photographs of the faces Tsimane' men and of male Andean herders/horticulturalists from small villages in Salta, Argentina (Sell et al., 2009), findings that again parallel those from studies involving faces and observers from WEIRD populations (Holzleitner & Perrett, 2016; Sell et al., 2009).

Together, this body of work suggests that judgements of threat potential contain some degree of accuracy. In addition to being accurate, these judgements were highly reliable across observers in each of the studies. In other words, judgements made by one observer tended to be highly correlated with judgements made by other observers. This consensus, which is also found for many other types of social judgements (e.g., agreeableness, conscientiousness, extraversion, honesty, see Table 1 in Zebrowitz & Collins, 1997), suggests that observers are attending to similar cues of threat potential in the face. Because most of the studies involved faces posed in neutral expressions, these cues are likely based in static or structural features of the face (e.g., bone structure) rather than dynamic ones (e.g., those related to emotional expressions). Additionally, because many of these findings persisted even when the researchers controlled for metrics of body size (e.g., weight, height, body mass index, Fink et al., 2007; Holzleitner & Perrett, 2016;

Sell et al., 2009; Trebicky et al., 2013), these structural features of the face may be revealing something about threat potential (e.g., propensity to harm) that is independent of traits advertised by the body (e.g., ability to harm), possibly explaining why the human visual system is particularly sensitive to faces when a stranger is first identified in a novel visual scene (Fletcher-Watson et al., 2008).

If humans are attending to static or structural cues in the face when judging threat potential, then researchers should be able to identify these specific cues and systematically examine their ecological validity (i.e., the associations between the cues and actual threat potential) and the extent to which they are utilized by observers (i.e., the associations between the cues and observers' judgements of threat potential) [as in Brunswick's (1955) Lens Model, reviewed in Zebrowitz & Collins, 1997]. Many of the studies above have identified some features, but they were often situated in the lower face (e.g., chin and lower jaw shape and size), a region that may be relatively masked when men are bearded, thus limiting the efficiency of this advertisement system across diverse populations. Further, participants appear better able to extract emotional information about threat from the upper than from the lower face (Bassili, 1979), which may speak to the greater importance of upper than lower face features for assessments of threat potential.

The Identification of a New Facial Metric: The Facial Width-to-Height Ratio

One recently identified metric that is located in the upper face and may be a part of an advertisement and assessment system is the facial width-to-height ratio (face ratio). First identified by Weston, Friday, and Liò (2007) in a set of dried skulls, this metric represents the distance between the left and right zygion (bizygomatic width) divided by the distance between the mid-brow and the upper lip (upper face height) (see Figure 5-1, Chapter 5). The face ratio was found to diverge in size for men and women at puberty (male > female), with this difference in the size of the face ratio not being explained by differences in the size of male and female bodies, indicating that selection pressures separate from those on the body may have acted on this facial metric (Weston et al., 2007)[4].

Because it is situated in the upper face, where humans extract information about threat (Bassili, 1979), and appeared to be sexually dimorphic, with this difference existing independent of body size, Carré and McCormick (2008) reasoned that this cue may represent one channel through which humans advertise and assess threat potential. In a series of studies, Carré and McCormick (2008) provided the first investigation of the ecological validity of the face ratio as a marker of threat potential. In the first study, the authors examined participants' aggressive behaviour in the laboratory using the Point Subtraction Aggression Paradigm (PSAP, first designed by Cherek, 1981), a well-validated behavioural measure of aggression (reviewed in Geniole, MacDonell, & McCormick, 2016). The authors found that individuals with larger face ratios were more aggressive on the task than were those with smaller face ratios, but the effects were specific to men. Within women, variation in the face ratio was not associated with variation in aggressive behaviour. In addition to these sex-specific associations, the size of the face ratio also differed between men and women, with men having larger face ratios on average than women, replicating the sex difference found by Weston and

[4] The sex difference resulted from male faces being much shorter than would be expected based on body size (Weston et al., 2007).

colleagues (2007) and extending these findings using measurements obtained from photographs of faces rather than from dried skulls.

In a second and third study, the authors investigated aggressive behaviour occurring outside of the laboratory, during university and professional hockey games. Aggressive behaviour in these studies was indexed by the average number of penalty minutes the men had per game (penalty minutes often result from behaviours intended to harm the other player, e.g., hitting, slashing, fighting). In both studies, men with larger face ratios had more penalty minutes per game than did men with smaller face ratios. Therefore, across three studies, men (but not women) with larger face ratios were more aggressive than were men with smaller face ratios, providing initial evidence that the face ratio may be an ecologically valid cue of threat potential.

To serve a social function, however, advertisements of threat potential must be perceivable and utilized by observers. In several follow-up studies, McCormick and colleagues investigated the extent to which observers' judgements of the photographs of the faces of the men from the PSAP study were correlated with men's face ratios. Consistent with the idea that the face ratio advertises threat potential, participants judged men with larger face ratios as more aggressive than they judged men with smaller face ratios (Carré, McCormick, & Mondloch, 2009). Further, these judgements of aggression were accurate: men who were judged as looking more aggressive had higher aggression scores during the PSAP than were men who were judged as less aggressive. These relationships between the face ratio and judgements of aggression persisted when photographs of the faces were presented rapidly (39 ms) or when blurred, which disrupts the processing of individual features but maintains the face ratio (Carré et al., 2009;

Carré, Morrissey, Mondloch, & McCormick, 2010), and when the photographs were cropped to remove forehead and lower face cues (Carré et al., 2010), which have been shown to influence judgements of dominance and masculinity (e.g., Keating, 1985; Windhager et al., 2011). Further, these relationships persisted when judgements of masculinity and attractiveness were controlled statistically, suggesting that the face ratio advertises threat potential independent of any effects of facial masculinity and attractiveness (Geniole, Keyes, Mondloch, Carré, & McCormick, 2012). These positive associations between the face ratio and judgements of aggression were also found when Chinese 8-year-old and young adult observers rated Caucasian faces and when Caucasian 8-year-old and young adult observers rated Chinese faces (Short et al., 2012), suggesting that observers may begin utilizing the face ratio to assess threat potential early in development. Across observers, perceptions of aggression also were linked more consistently to the face ratio of male than to female faces (Geniole et al., 2012), suggesting that this metric may be particularly relevant to the assessment of male threat potential. Therefore, observers appear to utilize the face ratio when judging aggressiveness, especially the aggressiveness of men, with these judgements forming rapidly and independently of experience with a face of a given race and of other facial features related to an individual's perceived attractiveness and masculinity. Together, the findings from these studies suggest that the face ratio may be an ecologically valid and utilized cue of male aggressiveness[5].

[5] The face ratio may be more relevant to advertisements and assessments of threat potential in men than in women because of men's greater threat potential (e.g., men are both physically stronger, and more physically aggressive than are women, reviewed in Archer, 2009; Sell et al., 2012), which would make attending to advertisements of threat more important in male than in female faces. On the other hand, women may more often express their aggressiveness through non-physical behaviours (e.g., gossip, social

Researchers from other labs, however, reported associations between this metric and behaviour aimed at exploiting others or cheating for personal gain (e.g., Haselhuhn & Wong, 2012; Stirrat & Perrett, 2010). For example, in Study 2 of Haselhuhn and Wong (2012), participants were asked to complete a variety of online self-report questionnaires and then, at the end of the study, to complete a dice role on another website (www.random.org/dice). Participants were informed that the value of their first dice roll would correspond to the number of ballots that would be entered, in their name, into a lottery for a cash prize. Thus, participants who rolled higher values would have better chances of winning the prize than participants who rolled lower values. After receiving the instructions, participants performed the dice rolls and reported the corresponding values. Importantly, the participants were able to report any value, as there was no validation procedure to ensure that the reported values corresponded to their first dice roll. This procedure provided participants the opportunity to cheat, undetected. Because dice rolls are random, any systematic association between the reported dice roll values and the face ratio of participants would indicate cheating behaviour. The researchers found that men (but not women) with larger face ratios reported higher dice roll values than men with smaller face ratios. Further, men's self-reported sense of power (example items: "I can get people to listen to what I say"; "I have a great deal of power"; scale designed by Anderson & Galinsky, 2006) was associated positively with both the male face ratio and the men's reported dice roll values, and partially accounted for the association between the face ratio and dice roll values, suggesting that personality traits related to the face ratio may better account for variability in cheating behaviours than this

ostracism) that have not been thoroughly investigated in relation to the face ratio. In the General Discussion, I return to sex differences in this metric, and its role in advertising and assessing threat.

facial feature itself. Therefore, rather than advertising aggressiveness specifically, this study and others (e.g., Stirrat & Perrett, 2010) suggested that the face ratio may advertise one's propensity to commit a variety of behaviours related more generally to threat potential.

Overview and scope of thesis studies

The overarching goal of this thesis was to better characterize the information carried by the face ratio and, in doing so, determine whether this metric and sensitivity to it may be part of an evolved advertisement and assessment system of threat potential. This research was guided by an Advertisement, Assessment, and Action model, through which I propose that the face ratio advertises, and can be used to assess, threat potential. This assessment, in turn, regulates the actions (e.g., approach/avoid, fight/flight) of observers. Although I do not test the underlying factors that may explain the face ratio's association with actual threat behaviour, I propose that this relationship arises because of underlying biological factors that simultaneously shape the structure of the face and the neural circuits that underlie personality traits and social behaviours (as in Carré & McCormick, 2008; also see "Path A" of the Developmental Model of Relationships Between Physical and Psychological Qualities, proposed by Zebrowitz & Collins, 1997). Because of this common underlying biological factor, the face ratio may sometimes be associated with behaviour, and was perhaps selected for in evolution because of its association with behaviour. In other words, I do not propose that the individual's face ratio influences their behaviour directly. Instead, I propose that the individual's face ratio is a correlate (and thus an advertisement) of the neural circuitry that underlies personality and social behaviour. Throughout Chapters 2-5, I use a variety of statistical techniques and terms (e.g., “mediation”, “explained”, “predicted”) to explain how an individual’s face ratio may be linked to their own behaviour. Some of these terms may imply that an individual’s face ratio directly influences their behaviour, or that their face ratio directly influences their personality. Such meaning is not intended[6]. These terms and analyses were used solely to express and test the idea that the individual’s face ratio is only associated with behaviour because of its association with neural circuitry, and thus I predict that measures more directly related to neural circuitry, such as personality traits, will share stronger associations with the individual’s behaviour, and that personality traits will also be correlated with the face ratio.

Based on the overarching goal of this thesis, and guided by the Advertisement, Assessment, and Action model, I conducted several studies across four chapters. In Chapter 2, I examine the extent to which the face ratio is associated with untrustworthy behaviour, using a more objective measure of cheating than that used by Haselhuhn and Wong (2012). In Chapter 3, I examine the extent to which observers utilize this metric for judgements of aggression and for judgements of trustworthiness. Although both judgement types may be indicators or behavioural expressions of a broader construct of “threat potential,” knowing which trait is more strongly and directly linked to the face ratio may provide insight into its specific social function and, thus, how observer sensitivity to the metric may have evolved. In Chapter 4, I reasoned that if the face ratio and sensitivity to it is part of an evolved advertisement and assessment system of threat potential, then associations between the face ratio and judgements of aggression should

[6] Through the Advertisement, Assessment, and Action model, I posit that the only variable the face ratio influences directly is the observer’s judgements, which in turn modulate the observer’s actions. Although I only investigate assessments (not actions) in this thesis, I discuss studies of observers’ actions in the General Discussion (Chapter 6).

persist even when men are bearded (the majority of studies on social perceptions involve the faces of clean-shaven men). Such a finding, for example, would indicate that this metric could have been operational in our ancestral past when many interactions likely involved bearded rather than shaved men. If the face ratio is not associated with judgements of aggression when men are bearded, it raises doubts about it belonging to an evolved system designed for advertising and assessing threat.

In Chapter 5, I provide a synthesis of all studies published on the face ratio before December 31, 2014 by performing several meta-analyses. These meta-analyses allowed me to investigate the metric's overall ecological validity (i.e., associations between the face ratio and a variety of behaviours related to threat potential) and the extent to which observers utilize the metric for judgements (i.e., the association between the face ratio and judgements related to threat potential). These meta-analyses provide the most precise estimates of the relationships between the face ratio and behaviour and between the face ratio and social judgements, using findings from studies involving participants from many different countries around the world (e.g., Canada, USA, Mexico, Germany, Czech Republic, China, Japan, countries within the United Kingdom). Based on the hypothesis that this metric and sensitivity to it is part of an evolved advertisement and assessment system of threat potential, I predicted that there would be significant associations between the face ratio and behaviours related to threat potential, and between the face ratio and judgements related to threat, when pooled across these diverse studies.

References

Anderson, C., & Galinsky, A. D. (2006). Power, optimism, and risk-taking. *European Journal of Social Psychology*, *36*(4), 511–536.

Archer, J. (2006). Testosterone and human aggression: an evaluation of the challenge hypothesis. *Neuroscience & Biobehavioral Reviews*, *30*(3), 319–345.

Archer, J. (2009). Does sexual selection explain human sex differences in aggression? *Behavioral and Brain Sciences*, *32*(3-4), 249–66; discussion 266–311.

Arnott, G., & Elwood, R. W. (2009). Assessment of fighting ability in animal contests. *Animal Behaviour*, *77*(5), 991–1004.

Aviezer, H., Trope, Y., & Todorov, A. (2012). Body Cues, Not Facial Expressions, Discriminate Between Intense Positive and Negative Emotions. *Science*, *338*, 1225–1229.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion*, *6*(2), 269–78.

Bassili, J. N. (1979). Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology*, *37*(11), 2049–58.

Berglund, A., & Rosenqvist, G. (2009). An intimidating ornament in a female pipefish. *Behavioral Ecology*, *20*(1), 54–59.

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

Berry, D. S. (1991). Accuracy in Social Perception: Contributions of Facial and Vocal Information. *Journal of Personality and Social Psychology*, *61*(2), 298–307.

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*(4), 991–998.

Bro-Jørgensen, J., & Beeston, J. (2015). Multimodal signalling in an antelope: fluctuating facemasks and knee-clicks reveal the social status of eland bulls. *Animal Behaviour*, *102*, 231–239.

Bruce, V., & Young, A. (2012). *Face Perception*. New York, NY: Psychology Press.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*(3), 193–217.

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 of the Royal Society B Biological Sciences*, *275*(1651), 2651–6.

Carré, J. M., McCormick, C. M., & Hariri, A. R. (2011). The social neuroendocrinology of human aggression. *Psychoneuroendocrinology*, *36*(7), 935–944.

Carré, J. M., McCormick, C. M., & Mondloch, C. J. (2009). Facial structure is a reliable cue of aggressive behavior. *Psychological Science*, *20*(10), 1194–1198.

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.

Carré, J. M., & Olmstead, N. A. (2015). Social neuroendocrinology of human aggression: Examining the role of competition-induced testosterone dynamics. *Neuroscience*, *286*, 171–186.

Cherek, D. R. (1981). Psychopharmacology effects of smoking different doses of nicotine on human aggressive behavior. *Psychopharmacology*, *75*, 339–345.

Cherulnik, P. D., Way, J. H., Ames, S., & Hutto, D. B. (1981). Impressions of high and low Machiavellian men. *Journal of Personality*, *49*(4), 388–400.

Dabbs, J. M. (1997). Testosterone, smiling, and facial appearance. *Journal of Nonverbal Behavior, 21*, 45–55.

Daly, M., & Wilson, M. (1988). *Homicide*. (S. B. Hrdy & R. W. Wrangham, Eds.). New York, NY: Aldine de Gruyter.

Eisner, M. P. (2003). Long-term historical trends in violent crime. *Crime and Justice, 30*, 83–142.

Ekman, P. (2003). Darwin, deception, and facial expression. *Annals of the New York Academy of Sciences*, *1000*, 205–221.

Fink, B., Neave, N., & Seydel, H. (2007). Male Facial Appearance Signals Physical Strength to Women. *American Journal of Human Biology*, *19*, 82–87.

Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception*, *37*(4), 571–583.

Fridlund, B. A. (1994). *Human Facial Expression: An Evolutionary View*. San Diego, CA: Academic Press.

Gallup, A. C., O'Brien, D. T., White, D. D., & Wilson, D. S. (2010). Handgrip strength and socially dominant behavior in male adolescents. *Evolutionary Psychology*, *8*(2), 229–243.

Gallup, A. C., White, D. D., & Gallup, G. G. (2007). Handgrip strength predicts sexual behavior, body morphology, and aggression in male college students. *Evolution and Human Behavior*, *28*(6), 423–429.

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*(1), e30366.

Geniole, S. N., MacDonell, E. T., & McCormick, C. M. (2016). The Point Subtraction Aggression Paradigm as a laboratory tool for investigating the neuroendocrinology of aggression and competition. *Hormones and Behavior*. doi: 10.1016/j.yhbeh.2016.04.006

Haselhuhn, M. P., & Wong, E. M. (2012). Bad to the bone: facial structure predicts unethical behaviour. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 571–576.

Hawley, P. H., & Vaughn, B. E. (2003). Aggression and Adaptive Functioning: The Bright Side to Bad Behavior. *Merrill-Palmer Quarterly*, *49*(3), 239–242.

Hehman, E., Flake, J. K., & Freeman, J. B. (2015). Static and dynamic facial cues differentially affect the consistency of social evaluations. *Personality and Social Psychology Bulletin*, *41*, 1123– 1134.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, *33*(2-3), 61–83; discussion 83–135.

Holtzman, N. S. (2011). Facing a psychopath: Detecting the dark triad from emotionally-neutral faces, using prototypes from the Personality Faceaurus. *Journal of Research in Personality*, *45*(6), 648–654.

Holzleitner, I. J., & Perrett, D. I. (2016). Perception of strength from 3D faces is linked to facial cues of physique. *Evolution and Human Behavior*, *37,* 217-229.

Järvi, T., & Bakken, M. (1984). The function of the variation in the breast stripe of the great tit (Parus major). *Animal Behaviour*, *32*(2), 590–596.

Keating, C. F. (1985). Gender and the physiognomy of dominance and attractiveness.

Social Psychology Quarterly, *48*(1), 61–70.

Kwon, M., Setoodehnia, M., Baek, J., Luck, S. J., & Oakes, L. M. (2016). The development of visual search in infancy: Attention to faces versus salience. *Developmental Psychology*, *52*(4), 537–555.

Marsh, A. A., Adams, R. B., & Kleck, R. E. (2005). Why do fear and anger look the way they do? Form and social function in facial expressions. *Personality & Social Psychology Bulletin*, *31*(1), 73–86.

Marsh, A. A., Cardinale, E. M., Chentsova-Dutton, Y. E., Grossman, M. R., & Krumpos, K. A. (2014). Power plays: Expressive mimicry of valid agonistic cues. *Social Psychological and Personality Science*, *5*(6), 684–690.

Maynard Smith, J., & Harper, D. (2003). *Animal signals*. New York, NY: Oxford University Press.

Mazur, A., & Booth, A. (1998). Testosterone and dominance in men. *The Behavioral and Brain Sciences*, *21*(3), 353–363.

McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review*, *90*(3), 215–238.

McGugin, R., & Gauthier, I. (2013). Face Recognition. In K. Ochsner & S. M. Kosslyn (Eds.), *The Oxford Handbook of Cognitive Neuroscience*, *Volume 2: The Cutting Edges* (p. 165). Oxford University Press.

Mueller, U., & Mazur, A. (1996). Facial Dominae of West Point Cadets as a Predictor of Later Military Rank. *Social Forces*, *74*(3), 823–850.

Muñoz-Reyes, J. A., Gil-Burmann, C., Fink, B., & Turiegano, E. (2012). Physical strength, fighting ability, and aggressiveness in adolescents. *American Journal of*

Human Biology, *24*(5), 611–617.

Nelson, N. L., & Russell, J. A. (2013). Universality Revisited. *Emotion Review*, *5*(1), 8–15.

Puts, D. A., Apicella, C. L., & Cárdenas, R. A. (2012). Masculine voices signal men's threat potential in forager and industrial societies. *Source: Proceedings: Biological Sciences*, *279*(1728), 601–609.

Scott-Phillips, T. C. (2008). Defining biological communication. *Journal of Evolutionary Biology*, *21*(2), 387–395.

Sell, A., Bryant, G. A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., … Gurven, M. (2010). Adaptations in humans for assessing physical strength from the voice. *Proceedings of the Royal Society B Biological Sciences*, *277*(1699), 3509–3518.

Sell, A., Cosmides, L., & Tooby, J. (2014). The human anger face evolved to enhance cues of strength. *Evolution and Human Behavior*, *35*(5), 425–429.

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 of the Royal Society B: Biological Sciences*, *276*(1656), 575–584.

Sell, A., Hone, L. S. E., & Pound, N. (2012). The importance of physical strength to human males. *Human Nature*, *23*(1), 30–44.

Setchell, J. M., Smith, T., Wickings, E. J., & Knapp, L. a. (2008). Social correlates of testosterone and ornamentation in male mandrills. *Hormones and Behavior*, *54*(3), 365–372.

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*(2), 121–129.

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*(4), 298–303.

Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: male facial width and trustworthiness. *Psychological Science*, *21*(3), 349–54.

Tibbetts, E. A., & Dale, J. (2004). A socially enforced signal of quality in a paper wasp. *Nature*, *432*, 218–223.

Tibbetts, E. A., & Lindsay, R. (2008). Visual signals of status and rival assessment in Polistes dominulus paper wasps. *Biology Letters*, *4*(3), 237–239.

Toscano, H., Schubert, T. W., & Sell, A. N. (2014). Judgments of dominance from the face track physical strength. *Evolutionary Psychology*, *12*(1), 1–18.

Trebicky, V., Havlícek, J., Roberts, S. C., Little, A. C., & Kleisner, K. (2013). Perceived aggressiveness predicts fighting performance in mixed-martial-arts fighters. *Psychological Science*, *24*(9), 1664–1672.

Undurraga, E. A., Eisenberg, D. T. A., Magvanjav, O., Wang, R., Leonard, W. R., McDade, T. W., … Godoy, R. A. (2010). Human's cognitive ability to assess facial cues from photographs: a study of sexual selection in the Bolivian Amazon. *PloS One*, *5*(6), e11027.

Walker, P. L. (2001). A Bioarchaeological Perspective on the History of Violence. *Annual Review of Anthropology*, *30*, 573–596.

Weston, E. M., Friday, A. E., & Liò, P. (2007). Biometric evidence that sexual selection has shaped the hominin face. *PloS One*, *2*(8), e710.

Windhager, S., Schaefer, K., & Fink, B. (2011). Geometric morphometrics of male facial shape in relation to physical strength and perceived attractiveness, dominance, and masculinity. *American Journal of Human Biology*, *23*(6), 805–814.

World Health Organization. (2002). *World report on violence and health : summary*. Geneva, Switzerland.

Yik, M. S. M., & Russell, J. A. (1999). Interpretation of Faces : A Cross-cultural Study of a Prediction from Fridlund ' s Theory. *Cognition and Emotion*, *13*(1), 93–104.

Zebrowitz, L. A., & Collins, M. A. (1997). Accurate social perception at zero acquaintance: the affordances of a Gibsonian approach. *Personality and Social Psychology Review*, *1*(3), 204–223.

Rationale for Chapter 2

In this chapter, I examine whether the face ratio is associated with untrustworthy behaviour using a more objective measure of cheating than that used by Haselhuhn and Wong (2012). Rather than examining correlations between men's face ratios and their self-reported dice roll values as an indirect index of cheating, I used a hidden screen recorder to capture the participants' actual dice roll value(s). Then, at the end of the study, I compared the values obtained from each participant's first dice roll with the number of ballots each participant entered into the lottery box. Doing so allowed for a direct measure of cheating behaviour for each participant, reducing any potential error variability that may have obscured the results of Haselhuhn and Wong (2012) (e.g., it is possible that some men with larger face ratios from their study did roll higher values in the study, and reported the values honestly). Further, I extend the findings of Haselhuhn and Wong (2012) by examining another set of personality traits that may be more strongly linked with the face ratio and threat potential, psychopathic personality traits. The personality measure I used in Chapter 2 not only captures some of the same "sense of power" traits investigated by Haselhuhn and Wong (2012), but has also been more extensively and psychometrically validated than has the scale used by Haselhuhn and Wong (2012).

Chapter 2: Fearless dominance mediates the relationship between the facial width-to-height ratio and cheating

The published version of this chapter is:

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

Author Contribution: I was the primary investigator of this work, responsible for the majority of the study design, data analysis, and writing of the manuscript.

Introduction

The facial width-to-height ratio (fWHR; bizygomatic width divided by upper-face height), first described by Weston, Friday, and Liò (2007), has garnered much attention because of its association with a cluster of behavioural tendencies in men, but not in women. For example, men with larger face ratios were more aggressive on a laboratory aggression measure than were men with smaller face ratios (Carré & McCormick 2008) and violent !Kung San men of Namibia had wider faces than those who were non-violent (Christiansen & Winkler, 1992). Amygdala activation, which predicts aggression in clinical populations (reviewed in Coccaro, Sripada, Yanowitch, & Phan, 2011), shared stronger associations with self-reported aggression in men with larger than with smaller face ratios (Carré, Murphy, & Hariri, 2013). Men with larger face ratios were also more likely to exploit the trust of others for personal gain (Stirrat & Perrett, 2010, 2012), endorse prejudicial beliefs (Hehman, Leitner, Deegan, & Gaertner, 2013), use explicit deception, and cheat in a lottery for a cash prize (Haselhuhn & Wong, 2012) than were men with smaller face ratios; these relationships were absent for women. Further, elite hockey players with larger face ratios had more penalty minutes per game (e.g., slashing, elbowing) than those with smaller ratios (Carré & McCormick, 2008). Although this association was only marginally significant in a larger sample of players ($p = 0.057$; Deaner, Goetz, Shattuck, & Schnotala, 2012), it appears to be moderated such that it is stronger among men who are lower in social status (Goetz et al., in press).

Any relationship between fWHR and such antisocial behaviour likely involves psychological mechanisms. Nevertheless, we have not found relationships between fWHR and broad domains of personality (e.g., such as the “big five” personality traits) (unpublished observations). Targeting specific personality traits rather than broad dimensions may be more fruitful. One study reported a correlation between fWHR and self-ratings of psychological "sense of power" in men, and that sense of power mediated the relationship between fWHR and cheating (Haselhuhn &Wong, 2012). The "sense of power" scale (Anderson & Galinsky, 2006), however, has not received the extensive psychometric analyses conducted for other questionnaires. Further, rather than directly measure cheating, Haselhuhn and Wong (2012) asked participants (50 men, 53 women) to report dice roll values (which were exchangeable for lottery ballots). Men, but not women, with high fWHRs reported higher dice rolls than those with low fWHRs, which the researchers concluded indicated cheating. Thus, cheaters and non-cheaters could not be directly compared to determine the true effect size. To address this limitation, we measured cheating directly in a larger sample (146 men and 76 women). We also used a measure of targeted personality traits that may better account for variability in fWHR and in cheating behaviour (and in antisocial behaviour, more generally). We measured psychopathic personality traits because of their robust association with antisocial behaviour in clinical and community samples (reviewed in Leistico, Salekin, DeCoster, & Rogers, 2008; Reidy, Shelley-Tremblay, & Lilienfeld, 2011).

We used the well-validated Psychopathic Personality Inventory-Revised (PPI-R; Lilienfeld & Widows, 2005), which assesses multiple personality traits relevant to psychopathy that load onto three factors: fearless dominance (low anxiety/stress, fearlessness, and high social dominance and influence), self-centred impulsivity (tendency to exploit others and to blame others for personal failures, and impulsivity), and coldheartedness (tendency to be apathetic, guiltless, and callous). We hypothesized that fearless dominance would be most relevant to cheating on the basis that it predicted antisocial behaviour (self-benefiting/other-costing behaviour) in versions of a Dictator game (Geniole, Busseri, & McCormick, under review) and because this factor contains items similar to the "sense of power" scale (e.g., fearless dominance: I am good at getting people to do favors for me, I often end up being the leader of a group, I have an easy time standing up for my rights; sense of power: I can get others to do what I want; If I want to, I get to make the decisions; My ideas and opinions are rarely ignored)[7]. Nevertheless, fearless dominance is distinct from sense of power in that it includes items that assess fearlessness and stress immunity. These characteristics may increase cheating by reducing the fears associated with being identified as a cheater. Furthermore, fearless dominance was related positively to achievement drive (Benning, Patrick, Hicks, Blonigen, & Krueger, 2003) and sensation seeking (reviewed in Poythress & Hall, 2011), traits that promote cheating (e.g., Williams, Nathanson, & Paulhus, 2010; DeAndrea, Carpenter, Shulman, & Levine, 2009). Therefore, we predicted that fWHR in men would be associated with cheating and psychopathic personality traits (specifically fearless dominance), that cheaters would be higher in fearless dominance than non-cheaters, and that the relationship between fWHR and cheating would be mediated by fearless dominance. We examined relationships among women, but predicted that the associations between fWHR, cheating, and personality would be specific to men.

Methods

Participants

[7] Some of the items were originally reverse keyed; I paraphrased them here to be directionally consistent with the trait.

Procedures were approved by the universities' (Brock and Wayne State) Research Ethics Board. Participants were recruited through online research pools (146 men and 77 women, M_{age} = 20.28, $SD._{age}$ = 2.79, 67% White, 5% Asian, 11% Black, 17% Other) and consented to the procedures. One participant was removed because her hijab limited facial measurements.

Measure of Psychopathic Personality Traits

The 154 item Psychopathic Personality Inventory – Revised (PPI-R; Lilienfeld & Widows, 2005) includes eight content scales subsumed by three factors: Fearless Dominance, Self-Centred Impulsivity, and Coldheartedness (described in Section 1). The factors are internally consistent (Cronbach's $\alpha \geq .78$ for each factor scale), possess high test-retest reliability (*r*s $\geq .82$; Lilienfeld & Widows, 2005), and are correlated with other self-report measures of psychopathic personality traits (Marcus, Fulton, & Edens, 2012). In the current sample, Fearless Dominance ($\alpha = .91$), Self-Centred Impulsivity ($\alpha = .89$), and Coldheartedness ($\alpha = .80$) were also internally consistent. The PPI-R questionnaire was embedded into a set of unrelated tasks administered for a different study.

Measure of Cheating

To measure cheating, we modified the dice rolling/lottery procedure of Haselhuhn and Wong (2012) in which the number of ballots a participant could enter into a lottery was determined by a dice roll. After completing the test battery, participants were given blank ballots, the lottery-box (into which ballots would be entered), a pen, and a printout of instructions: "(1) Go to www.random.org/dice; (2) Click 'roll dice' once. This will roll a pair of dice; (3) Add the numbers on each die together. This will equal the number of ballots you can enter into the raffle."

Participants were told that because several participants were in different rooms, the researcher had to remain available in the hallway. This procedure provided the participant the opportunity to cheat (entering more ballots into the lottery-box than the value of the dice roll) “undetected.” Hidden software recorded participants’ computer activity during the dice rolling procedure.

Facial Width-to-Height Ratio (FWHR)

After the lottery, participants went to the hallway to be photographed posed in a neutral facial expression for measurement of fWHR according to landmarks described in Weston and colleagues (2007) as in our previous studies (e.g., Carré & McCormick, 2008). Research assistants (blind to hypotheses) measured height (distance between lip and brow) and width (distance between left and right zygion) using ImageJ (NIH software). Inter-rater reliability was high for width, height, and fWHR (*r*s > .87).

Statistical Analysis

To simplify interpretation of results, we conducted 2 x 2 analyses of variance to determine if fWHR or the three factors of psychopathy differed for men versus women and cheaters versus non-cheaters (point-biserial correlations among variables produce the same results). To determine whether fWHR was related to the psychopathic personality factors, we entered the three factors as simultaneous predictors of fWHR. Bootstrapped mediation analysis (Preacher & Hayes, 2008) also was conducted to determine if the relationship between fWHR and cheating was mediated by psychopathic personality factors. Although three cases were identified as influential on specific regression coefficients and as multivariate outliers, removal of these cases did not alter the results significantly. Thus, all cases were included in the analyses reported.

Results

fWHR and psychopathic personality traits as a function of Sex and Cheating

Men (13%) and women (20%) did not differ in the percent that cheated in the lottery, $\chi^2 = 1.74$, $p = 0.13$. A two-factor (Sex = men vs. women; Cheating = cheaters vs. non-cheaters) ANOVA on fearless dominance scores indicated a main effect of Sex (men > women: $F_{1,218} = 38.06$, $p < 0.001$) that was obviated by a significant interaction between Sex and Cheating ($F_{1,218} = 8.80$, $p < 0.01$; see Figure 2-1-A). Follow-up t-tests indicated that for men ($t_{144} = -2.55$, $p = 0.01$, Cohen's $d = .64$), but not for women ($t_{74} = 1.78$, $p = 0.08$, *Cohen's* $d = .48$), cheaters had higher fearless dominance scores than did non-cheaters (see Table 2-1). Men had higher coldheartedness ($F_{1,218} = 24.88$, $p < 0.001$, Cohen's $d = .92$; see Figure 2-1-B) and higher, but not statistically significant, self-centred impulsivity scores than did women ($F_{1,218} = 2.96$, $p = 0.09$; see Figure 2-1-C), but there was no main effect of Cheating or interaction of Sex and Cheating for either coldheartedness or self-centred impulsivity (ps > 0.25).

A two-factor ANOVA on fWHR indicated that there were no main effects (Fs < 1.67, ps > 0.19) or interaction ($F_{1,218} = 2.25$, $p = 0.14$). Nevertheless, we conducted follow-up t-tests for the sexes separately based on our prediction that the relationship between fWHR and cheating would be specific to, or driven by, men. Results indicated that within men ($t_{144} = -1.97$, $p = 0.05$, Cohen's $d = .49$), but not within women ($t_{74} = 0.17$, $p = 0.86$, Cohen's $d = .08$), cheaters had larger fWHRs than did non-cheaters (see Figure 2-1-D).

Is fWHR related to psychopathic personality traits?

In men, the linear regression predicting fWHR was significant ($F_{3,142} = 2.76$, $p = 0.05$, $R^2 = .06$); fearless dominance was the only significant predictor ($t = 2.05$, $\beta = .17$, $p = 0.04$; other *p*s > 0.05) (see Figure 2-2-A). In women, the model was significant ($F_{3,72} = 3.60$, $p = 0.02$, $R^2 = .13$) and self-centred impulsivity ($t = -2.02$, $\beta = -.23$, $p = 0.05$; see Figure 2-2-B) and coldheartedness ($t = 2.72$, $\beta = .31$, $p < 0.01$; see Figure 2-2-C) were significant predictors but fearless dominance was not ($t = 0.62$, $\beta = .07$, $p = .54$).

Does fearless dominance mediate the relationship between fWHR and cheating in men?

We used binary logistic regression to test whether the relationship between fWHR and cheating was mediated by fearless dominance in men. fWHR was entered as the independent variable (Step 1), fearless dominance as the mediator (Step 2), and cheating as the dependent variable. Consistent with our prediction, the association between fWHR and cheating ($B = 3.13$, $Wald = 3.71$, $p = 0.05$) became non-significant ($B = 2.63$, $Wald = 2.57$, $p = 0.11$) when fearless dominance ($B = 0.03$, $Wald = 4.92$, $p = 0.02$) was added to the model as an additional predictor (Overall Model: $\chi^2 = 9.09$, $p = 0.01$, $Nagelkerke\ R^2 = .11$).

We also tested the prediction using bootstrapped mediation analysis with 5000 random samplings of the data with replacement. The mediation analysis indicated that the association between fWHR and cheating was indirect ($B = 0.58$; 95% confidence interval lower = 0.02, upper = 1.75) and mediated by fearless dominance (confidence intervals that do not overlap with a value of zero are indicative of mediation, Preacher & Hayes, 2008).

In men, extent of cheating (number of extra ballots entered into lottery) was associated positively with fWHR ($r = .23$, $p < 0.01$) and fearless dominance ($r = .21$, $p = 0.01$), but not with the other personality factors (*p*s > 0.66). Further, fearless dominance ($\beta = .16$, $t = 1.98$, $p = 0.05$) and fWHR ($\beta = .18$, $t = 2.24$, $p = 0.03$) were independent predictors of extent of cheating in a linear regression analysis and the face ratio did not have a significant indirect effect on the extent of cheating in a bootstrapped mediation analysis ($B = 0.19$; 95% confidence interval lower = -0.001, upper = .73); therefore, fearless dominance partially mediated the link between the face ratio and willingness to but not extent of cheating. In women, extent of cheating was associated negatively with fearless dominance ($r = -.26$, $p = 0.03$) but not with fWHR or the other personality factors (*p*s > .21).

Discussion

We investigated the extent to which psychopathic personality traits were related to the facial width-to-height ratio (fWHR) and mediated the link between this metric and antisocial behaviour, specifically cheating in a lottery for a cash prize. Consistent with our predictions, men with larger fWHRs were more willing to cheat, cheated to a greater extent, and had higher scores on the psychopathic personality factor of fearless dominance. In men, fearless dominance mediated the relationship between the fWHR and willingness to cheat, but not extent of cheating. Thus, our results confirm and extend the previous report by Haselhuhn and Wong (2012), who found a relationship between estimated cheating and the fWHR in men to be mediated by psychological sense of power.

The finding that fWHR and fearless dominance were associated with cheating adds to evidence that the fWHR is associated with antisociality (e.g., aggression, Carré &

McCormick, 2008, Goetz et al., in press; deception, Haselhuhn & Wong, 2012; prejudice, Hehman, Leitner, Deegan, & Gaertner, 2013; exploiting the trust of others, Stirrat & Perrett, 2012). Such traits, however, may confer success. For example, fWHR was a positive predictor of achievement drive in presidents (Lewis, Lefevre, & Bates, 2012), financial success among chief executive officers (Wong, Ormiston, & Haselhuhn, 2011; see also Alrajih & Ward, 2013), homerun frequency in baseball players (Tsujimura & Banissy, 2013), and with reproductive success (Loehr & O'Hara, 2013; near significant effect also reported in Gómez-Valdés et al., 2013, $p = 0.053$). Further, the relationship between fWHR and competitive behaviour may be stronger towards out-group than in-group members; fWHR was associated with prosocial or self-sacrificing behaviour towards in-group members when competition with the out-group was salient (Stirrat & Perrett, 2012). Similarly, although fearless dominance is considered a psychopathic trait and associated with aggression (Birkley, Giancola, & Lance, 2012; Denson, White, & Warburton, 2009), high fearless dominance also predicted greater achievement, well-being, education level, and class rank (Benning et al., 2003). Many researchers have suggested that traits within this factor (i.e., fearlessness, superficial charm) may promote success in business and politics, whereas the maladaptive behaviours stem from other trait factors (reviewed in Hall & Benning, 2006). Thus, despite some correlations with behaviours that seem antisocial, properly channelled aggression and dominance associated with having a larger fWHR may confer benefits and positive outcomes.

Fearless dominance and fWHR might share a common biological origin. Male and female face shapes diverge during puberty coincident with rising testosterone concentrations (Marečková et al., 2011; Verdonck, Gaethofs, Carels, & de Zegher, 1999), which also shape neural circuitry (reviewed in Blakemore, 2012). Masculinity and/or dominance in the face is associated positively with testosterone concentrations (e.g., Marečková et al., 2011; Penton-Voak & Chen, 2004; Pound, Penton-Voak, & Surridge, 2009; Swaddle & Reierson, 2002), and a recent study reported positive associations between fWHR and testosterone concentrations (Lefevre, Lewis, Perrett, & Penke, 2013). Testosterone is associated with dominance, personalized power, leadership, and with antisocial behaviour and risk-taking (reviewed in Archer, 2006). Men with higher testosterone concentrations had greater concern about status and greater hormonal reactivity to threats to status (reviewed in Archer, 2006; Mehta & Josephs, 2010). If testosterone at puberty influences both fearless dominance and face structure, then relationships of both to each other and to situation-specific behaviours consistent with fearless dominant behaviour would be expected.

Whereas fearless dominance was associated positively with cheating in men, it was associated negatively with willingness to cheat (albeit marginally) and the extent of cheating in women. Further, women with larger fWHRs had higher scores on the factor of coldheartedness and lower scores on the factor of self-centred impulsivity than did women with smaller fWHRs, which was unexpected given the lack of relationships between fWHR and behaviour for women in previous studies (Carré & McCormick, 2008; Haselhuhn & Wong, 2012; Stirrat & Perrett, 2010). One study found that coldheartedness was associated positively with social exclusion behaviours, the use of malicious humour, and verbal aggression in university students (84 women, 19 men) (Warren & Clarbour, 2009). Such indirect, relational aggression may be more relevant for investigations of women than other measures of dominance and aggression, which are typically higher in men than in women (reviewed in Archer, 2009). Further, men typically score higher in psychopathic personality traits than do women (reviewed in Dolan & Völlm, 2009), and other studies have reported that sex moderates the associations between psychopathic personality traits and behaviours such as aggression (e.g., Miller & Lynam, 2003; Birkley et al., 2012) and cooperation in trust games (e.g., Rilling et al., 2007). Similarly, our finding that fearless dominance was associated positively with cheating in men and negatively with cheating in women may reflect that the construct of psychopathy leads to different behavioural manifestations in men and in women and that the construct itself may be qualitatively different between the sexes (e.g., Dolan & Völlm, 2009).

In conclusion, our findings add to evidence that fWHR may signal personality correlates underlying aggressive and untrustworthy behaviour. In keeping with the possibility that fWHR may be an "honest signal" of men's behavioural tendencies, observers' judgements of men's aggressive potential (Carré, McCormick, & Mondloch, 2009; Carré, Morrissey, Mondloch, & McCormick, 2010; Geniole, Keyes, Mondloch, Carré, & McCormick, 2012), dominance (Alrajih & Ward, 2013), intimidation (Hehman, Leitner, & Gaertner, 2013), trustworthiness (Efferson & Vogt, 2013; Kleisner, Priplatova, Frost, & Flegr, 2013; Stirrat & Perrett, 2010), and prejudice (Hehman, Leitner, Deegan, & Gaertner, 2013) were correlated positively with fWHR. Furthermore, correlations between observers' judgements of men's aggressive potential and fWHR were found irrespective of the sex and ethnicity of the observers (Chinese or Caucasian) and of the ethnicity of the men's faces (Short et al., 2012). Moreover, estimates of aggression were correlated with fWHR even when made by observers as young as eight (Short et al.,

2012) and as old as 90 years of age (Boshyan, Zebrowitz, Franklin, McCormick, & Carré, 2013).

There is evidence that the fWHR influences social interactions, with participants less likely to entrust money to men with larger fWHRs in trust games (Stirrat & Perrett, 2010). If the fWHR is indeed an honest signal of personality and behavioural propensities, a question to be answered is are there advantages to signalling one's aggressive and untrustworthy personality or are the advantages primarily in the perception of the signal?

Tables

Table 2-1.

Descriptive statistics [mean (*SD*)] for cheaters (19 men, 15 women) and non-cheaters (127 men, 61 women) split by sex.

	Men		Women	
	Cheaters	Non-Cheaters	Cheaters	Non-Cheaters
Face Ratio	1.86 (0.17)	1.78 (0.16)	1.79 (0.12)	1.80 (0.12)
Fearless Dominance	133.12 (18.24)	121.31 (18.86)	101.87 (19.42)	110.36 (15.79)
Self-Centred Impulsivity	145.32 (18.68)	149.24 (22.19)	142.80 (13.44)	137.87 (21.61)
Coldheartedness	34.42 (6.23)	34.16 (6.35)	28.07 (5.54)	28.25 (7.04)
Extent of cheating	3.58 (2.06)		2.93 (2.09)	

Figures

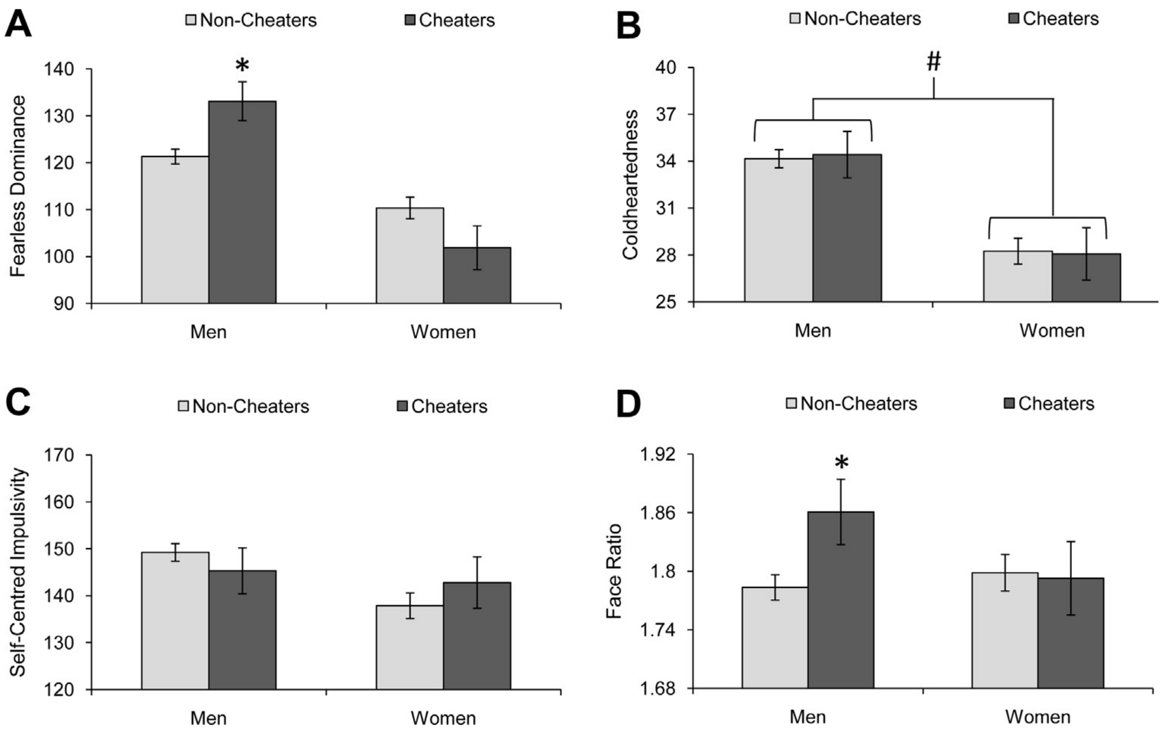

Figure 2-1. Mean (S.E.M.) fearless dominance (Panel A), coldheartedness (Panel B), self-centred impulsivity (Panel C), and fWHRs (Panel D) of participants who cheated (19 men, 15 women) or not (127 men, 61 women). * higher in cheaters, $p < 0.05$; # higher in men, $p < 0.05$.

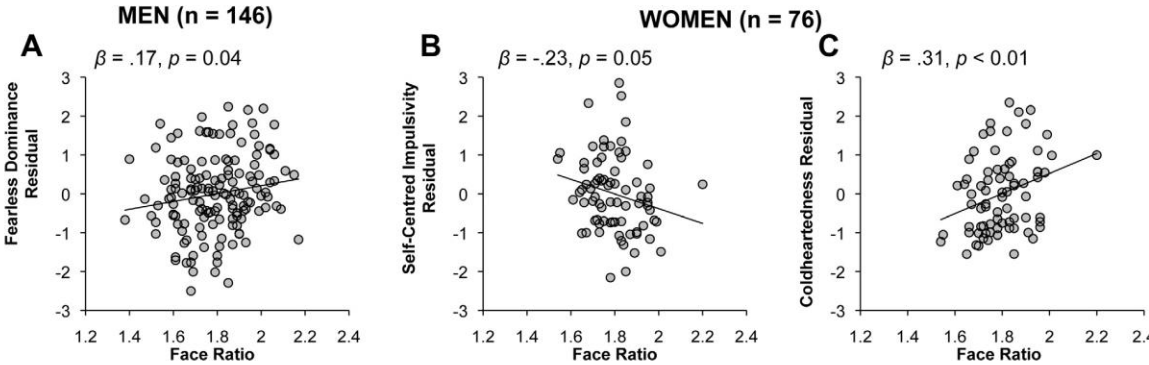

Figure 2-2. Scatterplots of the fWHR and fearless dominance residuals in men (Panel A), and of the fWHR and self-centred impulsivity residuals (Panel B) and coldheartedness residuals (Panel C) in women. Residuals were created by regressing each psychopathic personality factor on the other two psychopathic personality factors and saving the standardized residual.

References

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

Anderson, C., & Galinsky, A. D. (2006). Power, optimism, and risk-taking. *European Journal of Social Psychology, 36*, 511-536.

Archer, J. (2006). Testosterone and human aggression: An evaluation of the challenge hypothesis. *Neuroscience Biobehavioral Reviews, 30,* 319-345.

Archer, J. (2009). Does sexual selection explain human sex differences in aggression? *Behavioral and Brain Sciences, 32*, 249-266.

Benning, S. D., Patrick, C. J., Hicks, B. M., Blonigen, D. M., & Krueger, R. F. (2003). Factor structure of the psychopathic personality inventory: Validity and implications for clinical assessment. *Psychological Assessment*, *15*, 340-350.

Birkley, E. L., Giancola, P. R., & Lance, C. E. (2012). Psychopathy and the prediction of alcohol-related physical aggression: The roles of impulsive antisociality and fearless dominance. Drug and Alcohol Dependence, 128, 58-63

Blakemore, S. (2012). Imaging brain development: The adolescent brain. *NeuroImage, 61*, 397-406.

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*. doi: 10.1093/geronb/gbt054

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.

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 of the Royal Society B: Biological Sciences, 275*, 2651-2656.

Carré, J. M., McCormick, C. M., & Mondloch, C. J. (2009). Facial structure is a reliable cue of aggressive behavior. *Psychological Science, 20*, 1194-1198.

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.

Christiansen, K., & Winkler, E. (1992). Hormonal, anthropometrical, and behavioral correlates of physical aggression in !Kung San men of Namibia. *Aggressive Behavior, 18*, 271-280.

Coccaro, E. F., Sripada, C. S., Yanowitch, R. N., & Phan, K. L. (2011). Corticolimbic function in impulsive aggressive behavior. *Biological Psychiatry*, *69,* 1153-1159.

DeAndrea, D. C., Carpenter, C., Shulman, H., & Levine, T. R. (2009). The relationship between cheating behavior and sensation-seeking. *Personality and Individual Differences, 47*, 944-947.

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.

Denson, T. F., White, A. J., & Warburton, W. A. (2009). Trait displaced aggression and psychopathy differentially moderate the effects of acute alcohol intoxication and rumination on triggered displaced aggression. *Journal of Research in Personality*, *43*, 673-681.

Dolan, M., & Völlm, B. (2009). Antisocial personality disorder and psychopathy in women: A literature review on the reliability and validity of assessment instruments. *International Journal of Law and Psychiatry, 32*, 2-9.

Efferson, C., & Vogt, S. (2013). Viewing men's faces does not lead to accurate predictions of

trustworthiness. *Scientific Reports, 3,* 1047.

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.

Goetz, S. M. M., Shattuck, K. S., Miller, R. M., Campbell, J. A., Lozoya, E., Weisfeld, G. E., & Carré, J. M. (in press). Social status moderates the relationship between facial structure and aggression. *Psychological Science.*

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.

Hall, J. R., & Benning, S. D. (2006). The "successful" psychopath: Adaptive and subclinical manifestations of psychopathy in the general population. In C. J. Patrick (Ed.), *Handbook of psychopathy* (pp. 459-480). New York: The Guilford Press.

Haselhuhn, M. P., & Wong, E. M. (2012). Bad to the bone: Facial structure predicts unethical behaviour. *Proceedings of the Royal Society B: Biological Sciences, 279*, 571-576.

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.

Hehman, E., Leitner, J. B., & Gaertner, S. L. (2013). Enhancing static facial features increases intimidation. *Journal of Experimental Social Psychology*, *49,* 747-754.

Kleisner, K., Priplatova, L., Frost, P., & Flegr, J. (2013). Trustworthy-looking face meets brown eyes. *PloS one*, *8*, e53285.

Lefevre, C. E., Lewis, G. J., Perrett, D. I., & Penke, L. (2013). Telling facial metrics: facial width is associated with testosterone levels in men. *Evolution and Human Behavior*. doi: 10.1016/j.evolhumbehav.2013.03.005

Leistico, A. R., Salekin, R. T., DeCoster, J., & Rogers, R. (2008). A large-scale meta-analysis relating the Hare measures of psychopathy to antisocial conduct. *Law and Human Behavior, 32*, 28-45.

Lewis, G. J., Lefevre, C. E., & Bates, T. C. (2012). Facial width-to-height ratio predicts achievement drive in US presidents. *Personality and Individual Differences, 52*, 855-857.

Lilienfeld, S. O., & Widows, M. R. (2005). *Psychopathic personality Inventory–Revised: Professional manual.* Lutz, FL: Psychological Assessment Resources.

Loehr, J., & O'Hara, R. B. (2013). Facial morphology predicts male fitness and rank but not survival in Second World War Finnish soldiers. *Biology Letters*. doi: 10.1098/rsbl.2013.0049

Marcus, D. K., Fulton, J. J., & Edens, J. F. (2012). The two-factor model of psychopathic personality: Evidence from the psychopathic personality inventory. *Personality Disorders: Theory, Research, and Treatment, 3*, 140-154.

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.

Mehta, P. H., & Josephs, R. A. (2010). Testosterone and cortisol jointly regulate dominance: Evidence for a dual-hormone hypothesis. *Hormones and Behavior*, *58*, 898-906.

Miller, J. D., & Lynam, D. R. (2003). Psychopathy and the five-factor model of personality: A replication and extension. *Journal of Personality Assessment, 81*, 168-178.

Penton-Voak, I. S., & Chen, J. Y. (2004). High salivary testosterone is linked to masculine male facial appearance in humans. *Evolution and Human Behavior*, 25, 229–241.

Pound, N., Penton-Voak, I. S., & Surridge, A. K. (2009). Testosterone responses to competition in men are related to facial masculinity. *Proceedings of the Royal Society B: Biological Sciences, 276*, 153-159.

Poythress, N. G., & Hall, J. R. (2011). Psychopathy and impulsivity reconsidered. *Aggression and Violent Behavior, 16*, 120-134.

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.

Reidy, D. E., Shelley-Tremblay, J. F., & Lilienfeld, S. O. (2011). Psychopathy, reactive aggression, and precarious proclamations: A review of behavioral, cognitive, and biological research. *Aggression and Violent Behavior*, *16*, 512-524.

Rilling, J. K., Glenn, A. L., Jairam, M. R., Pagnoni, G., Goldsmith, D. R., Elfenbein, H. A., & Lilienfeld, S. O. (2007). Neural correlates of social cooperation and non-cooperation as a function of psychopathy. *Biological Psychiatry, 61*, 1260-1271.

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.

Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust. *Psychological Science, 21*, 349-354.

Stirrat, M., & Perrett, D. I. (2012). Face structure predicts cooperation. *Psychological Science*, *23*, 718-722.

Swaddle, J. P., & Reierson, G. W. (2002). Testosterone increases perceived dominance but not attractiveness in human males. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *269*, 2285-2289.

Tsujimura, H., & Banissy, M. J. (2013). Human face structure correlates with professional baseball performance: insights from professional Japanese baseball players. *Biology Letters*. doi: 10.1098/rsbl.2013.0140

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.

Warren, G. C., & Clarbour, J. (2009). Relationship between psychopathy and indirect aggression use in a noncriminal population. *Aggressive Behavior, 35*, 408-421. doi:10.1002/ab.20317

Weston, E. M., Friday, A. E., & Liò, P. (2007). Biometric evidence that sexual selection has shaped the hominin face. *PLoS ONE, 2*, e710.

Williams, K. M., Nathanson, C., & Paulhus, D. L. (2010). Identifying and profiling scholastic cheaters: Their personality, cognitive ability, and motivation. *Journal of Experimental Psychology: Applied, 16*, 293-307.

Wong, E. M., Ormiston, M. E., & Haselhuhn, M. P. (2011). A face only an investor could love: CEOs' facial structure predicts their firms' financial performance. *Psychological Science, 22*, 1478-1483.

Rationale for Chapter 3

In Chapter 2, I provided evidence that the face ratio is associated with cheating, adding to other studies that have reported associations between this metric and untrustworthy (cheating, deception, Haselhuhn & Wong, 2012; exploiting the trust of others for personal gain, Stirrat & Perrett, 2010) and aggressive (Carré & McCormick, 2008) behaviours. In Chapter 3, I examine the extent to which observers utilize the face ratio for assessing these two types of treat-related judgements (aggression versus untrustworthiness). Some studies provided evidence that observers utilize the face ratio for judgements of aggression (e.g., Boshyan, Zebrowitz, Franklin, McCormick, & Carré, 2013; Carré et al., 2009, 2010; Geniole et al., 2012; Lefevre & Lewis, 2013; Short et al., 2012) whereas others provided evidence that observers utilize the metric for judgements of trustworthiness (e.g., Efferson & Vogt, 2013; Kleisner, Priplatova, Frost, & Flegr, 2013; Stirrat & Perrett, 2010). Both judgements are likely indicators of a more general "threat potential" construct, but knowing which of the two is more directly linked to the face ratio may provide clues as to the social function of this metric and how perceptual systems evolved to become so sensitive to it. For example, if more strongly and directly linked to judgements of aggression, the face ratio may have primarily served to regulate decisions to fight or flee in conflict and thus visual systems may have evolved to be sensitive to the face ratio because such sensitivity allowed for judicious decision making in physical conflicts. If more strongly and directly linked to judgements of trustworthiness, then the face ratio may have primarily served to regulate decisions to trust or not in social and economic interactions and thus visual systems may have evolved to be sensitive to it because this sensitivity improved collaboration and cooperation decisions. Further, because judgements of aggression and of trustworthiness are highly negatively correlated (men judged as looking more aggressive are also judged as looking less trustworthy), it is possible that one of these judgements simply forms as a consequence of the other, and not directly because of the face ratio. I also tested this possibility across several studies in Chapter 3.

Chapter 3: The facial width-to-height ratio shares stronger links with judgements of aggression than with judgements of trustworthiness

The published version of this chapter is:

Geniole, S.N., Molnar, D.S., Carré, J.M., McCormick, C.M. (2014). The facial width-to-height ratio shares stronger links with judgements of aggression than with judgements of trustworthiness. *Journal of Experimental Psychology: Human Perception and Performance, 40*(40), 1526 – 1541. doi: 10.1037/a0036732

Author Contribution: I was the primary investigator of this work, responsible for the majority of the study design, data analysis, and writing of the manuscript.

Introduction

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 behaviour across numerous species (reviewed in Blanchard, Griebel, Pobbe, & Blanchard, 2011; Taylor & Elmwood, 2003). Further, static images are sufficient to enable accurate judgements; assessments of dominance in chimpanzees (Kramer, King, & Ward, 2011; Kramer & Ward, 2012) and of fighting ability in humans (Sell et al., 2009; Trebický, Havlícek, Roberts, Little, & Kleisner, 2013) based on facial and body photographs are above chance accuracy. Although information about another’s formidability and/or their behavioural intentions can be inferred from emotional expressions (Fridlund, 1994; McArther & 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 non-violent (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 (Trebický 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). 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 behaviour does not always predict actual behaviour 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 (mean age = 52.5 years), however, the face ratio was associated with dominance rather than aggression (Alrajih & Ward, 2013). Although judgements of masculinity and of aggression are highly correlated, the face ratio predicted judgements of aggression independently of judgements of masculinity (Geniole et al., 2012). Judgements 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' judgements of trust whereas minimizations increased observers' judgements 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 judgements. One study, however, found no association between the face ratio and behaviour in a trust game (Efferson & Vogt, 2013).

Overall, the face ratio is found to predict the extent to which men's behaviour is trustworthy or untrustworthy in addition to predicting aggressive behaviour.

Irrespective of the accuracy of social judgements based on viewing faces, there is ample evidence of the consistency of such judgements across observers and ample evidence that such judgements are relevant for social interactions (Todorov, Mende-Siedlecki, & Dotsch, 2013). There is much evidence that the face ratio is a cue in such judgements (as noted above), with associations between the face ratio and judgements of aggression/trustworthiness in men among the highest associations reported for judgements 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 judgements across individuals is notable in light of evidence that traits within the observer also influence judgements (e.g., Willis, Dodd, & Palermo, 2013). Determining the perceptual judgement 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 judgements.

With regard to the function and evolution of the face ratio, if variation in the face ratio is more strongly linked to judgements of aggression than to trustworthiness, then the face ratio may have functioned to reduce injury or the likelihood of death resulting from poor judgements 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 judgements 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 to 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 judgement 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 judgements of trustworthiness and aggression, and determining the order in which these judgements form, will also provide theoretical insight into the perceptual processes that shape such judgements. For example, although trustworthiness is considered a high level cognitive judgement, trust judgements form rapidly (Willis & Todorov, 2006) and seem to rely on subcortical brain structures such as the amygdala (e.g., Adolphs, Tranel, & Demasio, 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), judgements of trustworthiness may "piggyback" on other social judgements (threat, aggressiveness, dangerousness) that are better suited for such circumstances.

There is some evidence that judgements of aggression may be better suited for rapid assessments during first encounters with strangers than are judgements of trustworthiness. For example, judgements of aggression appear to be more accessible (or easier to assess), as they are provided at a frequency more than three times that of judgements of trustworthiness when describing faces (Oosterhof & Todorov, 2008). Further, although observers' confidence in trustworthiness judgements decreased with shortened exposure time to face stimuli, their judgements of aggression did not (Willis & Todorov, 2006). Thus, it is possible that judgements of aggression are more relevant for circumstances in which exposure to a face is limited or the history of the individual is unknown. As such, judgements of aggression may be a mechanism through which judgements of trustworthiness form in such circumstances. Nevertheless, no studies to date have tested this "piggyback" framework.

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

Study 1 Methods

Participants

Participants (29 women, 5 men; mean 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 (mean 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 variation in judgements related to ethnicity-based stereotypes) 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 (mean 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' judgements and/or obscure the relationships between the face ratio and these judgements (Kenny & Fletcher, 1973; de Souza et al., 2003).

Excluding individuals with facial hair reduced our sample to 25 men (mean 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 jewellery (e.g., earrings, necklace) were also modified using Adobe Photoshop to erase the visible jewellery and avoid observer bias in judgements.

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-dimensional facial modelling 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 3D 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 3-1). After the adjustment, the Facegen faces were saved as 8-bit greyscale, bitmap images. Greyscale was used to minimize the influence of colour tones in the face known to influence judgements 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: mean = 1.90, *SD* = 0.17; Facegen models: mean = 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' judgements 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 judgements share a similar factor structure if they are analysed 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. Inter-rater 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 inch Dell laptop monitor (approximately 15.2 x 12.9 visual degrees when viewed from 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 1000 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 versus Facegen) counterbalanced across participants. Within each block, a face was presented individually for 1000 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[8]; 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, no association. Linear regression was used to determine which judgements

---

[8] The aggression and trustworthiness blocks were not counterbalanced across participants; all participants rated aggression first, and trustworthiness second.

shared a unique association with the face ratio. Semi-partial 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 5000 samples of the aggregate data (Preacher & Hayes, 2008), and using multilevel modelling with HLM version 7.0 software (Raudenbush, Bryk, & Congdon, 2004). In multilevel modelling, every judgement is nested within each of the respective observers from which that judgement 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.

Study 1 Results

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

Original, clean-shaven faces (25 faces). Judgements of aggression were significantly and positively correlated with the face ratio for 16 of the 34 observers (mean ± C.I. = .34 ± .10; one-sample t-test: $t_{33} = 6.94$, $p < .001$). Judgements of trustworthiness for the original clean-shaven faces were significantly and negatively correlated with the face ratio for 13 of the 34 observers (mean ± C.I. = -.25 ± .11; one-sample t-test: $t_{33} = -4.57$, $p < .001$).

Facegen faces (54 faces). Judgements of aggression were significantly and positively correlated with the face ratio for 26 of the 33 observers (mean ± C.I. = .37 ±

.06; one-sample t-test: t_{32} = 11.88, p < .001). Judgements of trustworthiness were significantly and negatively correlated with the face ratio for 15 of the 33 (mean ± C.I. = -.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). Descriptive statistics are presented in Table 3-1. The face ratio was correlated positively 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 correlated negatively (see Figure 3-2-A). To determine whether the relationship between the face ratio and judgements 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 face ratio-trustworthiness relationship, we would expect that the association between the face ratio and judgements 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 3-2-A). Furthermore, adding ratings of aggression accounted for significantly more variability in the ratings of trustworthiness ($R^2 change$ = .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 et al., 1995; Mason et al., 1989; Neter et al., 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, β = .57, sr = .40, p = .03), but not trustworthiness (t = -0.24, β = - .06, sr = -.04, p = .81), was a significant predictor of variation in the face ratio (VIF = 2.10).

Analyses using a bootstrapping technique with 5000 samples (Preacher & Hayes, 2008), provided 95% confidence intervals (bias corrected) for the indirect relationship between the face ratio and judgements of trustworthiness (i.e., controlling statistically for judgements 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 judgements of trustworthiness (C.I. = -3.93, -0.79).

When the data were analysed using multilevel modelling, the average correlations between the face ratio and judgements of trustworthiness (γ = -2.38, s.e. = 0.47, t_{815} = -5.05, p < .001) and between the face ratio and judgements of aggression (γ = 3.01, s.e. = 0.46, t_{815} = 6.55, p < .001) were significant. The average association between judgements of aggression and judgements of trustworthiness was also significant (γ = -0.37, s.e. = 0.06, t_{815} = -6.52, p < .001). The average relationship between the face ratio and trustworthiness was attenuated, but nevertheless still significant (γ = -1.43 s.e. = 0.37, t_{814} = -3.89, p < .001) when judgements of aggression (γ = -.32, s.e. = 0.05, t_{814} = -5.91, p <

.001) was added as a simultaneous predictor, indicating partial mediation (see Table 3-2 for Sobel test).

Facegen faces (54 faces). See descriptive statistics in Table 3-1. The face ratio was correlated positively with the mean group ratings of aggression and negatively with the mean group ratings of trustworthiness; aggression and trustworthy ratings were correlated negatively (see Figure 3-2-B). 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 3-2-B). Further, ratings of aggression accounted for significantly more variability in the ratings of trustworthiness ($R^2 change$ = .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 judgements 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, β = .63, sr = .31, p < .01), but not trustworthiness (t = 0.14, β = .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 judgements of trustworthiness (C.I. = -2.84, -1.36). Using multilevel modelling, the average correlations between the face ratio and judgements of trustworthiness (γ = -1.93, s.e. = .30, t_{1748} = -6.54, p < .001) and between the face ratio and judgements of aggression (γ = 2.61, s.e. = 0.23, t_{1748} = 11.31, p < .001) were significant. The average association between judgements of aggression and of trustworthiness was also significant (γ = -0.35, s.e. = 0.04, t_{1748} = -8.38, p < .001). The average relationship between the face ratio and trustworthiness was attenuated but still significant (γ = -1.15, s.e. = 0.23, t_{1747} = -4.93, p < .001) when judgements of aggression (γ = -0.30, s.e. = 0.04, t_{1747} = -8.03, p < .001) was added as a simultaneous predictor, indicating partial mediation (see Table 3-2 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' judgements 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 face ratio-trustworthiness relationship in this new subset of faces.

Study 2 Methods

Participants

Participants (12 women, 12 men; mean age = 19.58, SD = 1.47, age range: 18-23 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.

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, second on aggression. After rating the faces on aggression, participants provided ratings of hairiness; "How much facial hair does this person have?" Responses were provided using a 7-point likert scale [1 = *some (a little)*, 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.

Study 2 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 using 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 judgements of aggression that were significantly correlated with the face ratio (mean ± C.I. = .23 ± .11; one-sample t-test: $t_{23} = 4.36$, $p < .001$) whereas only one of the 24 observers provided judgements of trustworthiness that were significantly associated with the face ratio (mean ± C.I. = -.12 ± .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?

See descriptive statistics in Table 3-1. When ratings were averaged across observers, judgements of aggression were correlated negatively with judgements of trustworthiness but the face ratio shared a significant association only with judgements of aggression (see Figure 3-2-C). 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 judgements of trustworthiness (albeit non-significant; p = .20), this relationship was completely attenuated when judgements of aggression were added to the model (see Figure 3-2-C). Further, judgements of aggression accounted for significantly more variability in the judgements of trustworthiness ($R^2 change$ = .32, p < .01) than did the face ratio alone (VIF = 1.26). When judgements 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, judgements of aggression approached significance (t = 1.73, β = .46, sr = .35, p = .10) but those of trustworthiness did not (t = 0.02, β = .01, sr = .00, p = .98) (VIF = 1.66).

Results from the bootstrapped mediation also indicated that the face ratio shared an indirect association with judgements of trustworthiness (C.I. = -2.20, -0.36). Using multilevel modelling, the average correlations between the face ratio and judgements of trustworthiness (γ = -1.06, s.e. = 0.35, t_{503} = -3.05, p = .002) and between the face ratio and judgements of aggression (γ = 2.33, s.e. = 0.53, t_{503} = 4.37, p < .001) were significant. The average association between judgements of aggression and trustworthiness was also significant (γ = -0.24, s.e. = 0.06, t_{503} = -3.99, p < .001). The average relationship between the face ratio and trustworthiness was attenuated (γ = -0.53, s.e. = 0.37, t_{502} = -1.43, p = .15) when judgements of aggression (γ = -0.23, s.e. = 0.06, t_{502} = -3.74, p < .001) was added as a simultaneous predictor, indicating mediation (see Table 3-2 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 a different 24 men than those used in the present studies.

Study 3 Methods

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 re-analysis 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 greyscale.

Procedure

Faces appeared for 2000 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 2. For more details regarding the procedures, see Carré and colleagues (2009).

Study 3 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 judgements of trustworthiness (t = 1.16, β = .13, sr = .11, p = .26) when judgements 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 ($R^2 change$ = .63, p < .001) than did the face ratio alone (VIF = 1.53). When both judgements were added as simultaneous predictors of the face ratio, ratings of aggression (t = 2.53, β = 1.00, sr = .43, p = .02) were a significant predictor of the face ratio but those of trustworthiness were not (t = 1.16, β = .46, sr = .20, p = .26) (VIF = 5.37).

Results from the bootstrapped mediation also indicated that the face ratio shared an indirect association with judgements of trustworthiness (C.I. = -5.35, -1.46). Using multilevel modelling, the average correlations between the face ratio and judgements of trustworthiness (γ = -2.41, s.e. = 0.40, t_{703} = -5.98, p < .001) and between the face ratio and judgements of aggression (γ = 4.03, s.e. = 0.25, t_{703} =15.87, p < .001) were significant. The average association between judgements of aggression and trustworthiness was also significant (γ = -0.40, s.e. = 0.03, t_{703} = -11.94, p < .001). The average relationship between the face ratio and trustworthiness was attenuated but still significant (γ = -0.92, s.e. = 0.44, t_{702} = -2.10, p = .04) when judgements of aggression (γ = -0.37, s.e. = 0.04, t_{702} = -9.46, p < .001) was added as a simultaneous predictor, indicating partial mediation (see Table 3-2 for Sobel test).

Study 4

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

Study 4 also had a couple of methodological advantages compared to Studies 1-3. Specifically, ratings were counterbalanced in Study 4 such that half of participants rated aggression first (and trustworthiness second) and the other half rated trustworthiness first (and aggression second). Although the consistency in our results from Studies 1-3, irrespective of the order of ratings (aggression rated first in Study 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 judgements of aggression and trustworthiness than did Studies 1-3 because the question used to prompt participants' judgements 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 judgements of trustworthiness ("How trustworthy does this person look?").

Study 4 Methods

Participants

Participants (32 women, 8 men; mean age = 19.38, *SD* = 1.86, age range: 17-26 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; mean 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. Inter-rater 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 Willis and Todorov (2006). All participants made judgements of both characteristics, with the order of the judgements counterbalanced across participants. Before providing these judgements, 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 judgements 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 centre of the computer screen, which had a black background. A white fixation-cross appeared in the centre 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 judgements 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 judgements, 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 ratings, participants read instructions about the next judgement they would provide and the same process was repeated.

Study 4 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: mean ms = 1048.71, *SD* = 66.05; trustworthiness first: mean ms = 1070.53, *SD* = 100.76; paired-sample t-test: $t_6 = 0.91$, $p = .40$; see Figure 3-3-A).

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 judgements, lengthy response times were Winsorized such that times longer than 4000 ms were changed to 4000 ms (greater than 2.5 standard deviations of the mean, 1.81% of aggression and 2.77% of trustworthiness response times)[9]. 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 outliers in many recent studies (e.g., Chambers, Swan, &

[9] The same results were obtained when lengthy response times (>4000 ms) were trimmed (i.e., removed from the analysis). Specifically, when data were trimmed, a mixed factorial ANOVA with one within-subjects factor (aggression versus trustworthiness judgements) 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 judgements 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 forty 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).

Heesacker, 2014; Wilkowski & Meier, 2010; Mueller et al., 2011; Townsend et al., 2014; Lai et al., 2012). A mixed factorial ANOVA with one within subjects factor (aggression versus trustworthiness judgements) 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 judgements faster if they rated aggression first rather than second ($F_{1, 128} = 9.03$, $p < .01$, Cohen's $d = 0.53$) (see Figure 3-3-B). 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 to the other, mean response times for aggression and for trustworthiness judgements were compared within each of the 40 observers using paired-samples t-tests (see Figure 3-3-C). 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 who rated aggression faster than trustworthiness was statistically significant ($\chi^2_1 = 7.12$, $p < .01$).

As another test of the hypothesis that judgements of aggression form faster than judgements 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 to 0 indicated that judgements of aggression were provided significantly faster than were judgements of trustworthiness (mean = 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.

Judgements of aggression were significantly and positively correlated with the face ratio for 24 of the 40 observers (mean ± C.I. = .26 ± .06; one-sample t-test: t_{39} = 9.58, $p < .001$). Judgements of trustworthiness for this same group were significantly and negatively correlated with the face ratio for 17 of the 40 observers (mean ± C.I. = -.21 ± .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?

See descriptive statistics in Table 3-1. The face ratio was correlated positively with the mean ratings of aggression and negatively with the mean ratings of trustworthiness; aggression and trustworthy ratings were correlated negatively (see Figure 3-2-D). 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 3-2-D). Further, ratings of aggression accounted for more variability in the ratings of trustworthiness ($R^2 change$ = .63, $p < .001$) than did the face ratio alone (VIF = 1.31). When both judgements were entered as simultaneous predictors of the face ratio, judgements of aggression were significant (aggression: t = 2.49, β = .60, sr = .28, p = .02) but judgements of trust were not (t = 0.55, β = .13, sr = .06, p = .58) (VIF = 4.77).

Results from the bootstrapped mediation also indicated that the face ratio shared an indirect association with judgements of trustworthiness (C.I. = -3.37, -1.43). Using multilevel modelling, the average correlations between the face ratio and judgements of trustworthiness (γ = -2.14, s.e. = 0.20, t_{2559} = -10.49, $p < .001$) and between the face ratio and judgements of aggression ($\gamma = 2.72$, s.e. = 0.28, $t_{2559}= 9.68$, $p < .001$) were significant. The average association between judgements of aggression and trustworthiness was also significant ($\gamma = -0.34$, s.e. = 0.04, $t_{2559} = -9.06$, $p < .001$). The average relationship between the face ratio and trustworthiness was attenuated, but still significant ($\gamma = -1.29$, s.e. = 0.25, $t_{2558} = -5.22$, $p < .001$) when judgements of aggression ($\gamma = -0.31$, s.e. = 0.04, $t_{2558} = -7.95$, $p < .001$) was added as a simultaneous predictor, again indicating partial mediation (see Table 3-2 for Sobel test)[10].

Did the order in which participants provided ratings influence how they judged the faces?

A mixed factorial ANOVA with one within subjects factor (aggression versus trustworthiness rating) and one between subjects factor (aggression rated first versus 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

---

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

before they provided judgements 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 rated circles as smaller (used lower values on the 7-point scale) than did those who rated aggression second (aggression first: mean = 4.61, *SD* = 2.05; aggression second: 4.80, *SD* = 1.90). Therefore, as opposed to order of judgements 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 judgements.

We also found that the associations between the face ratio and judgements of aggression (aggression rated first: $r = .40$, $p < .01$; aggression rated second: $r = .55$, $p < .001$) and between the face ratio and judgements 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 judgements: $z = -1.14$, $p = .25$; trustworthiness judgements: $z = 0.27$, $p = .78$). 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 judgements of aggression, of trustworthiness, or the relationship between both of these judgements and the face ratio[11].

Do judgements of aggression share stronger links with the face ratio than do judgements of trustworthiness?

To determine which judgement shared the strongest association with the face ratio, the associations between judgements of aggression and the face ratio and between judgements of trustworthiness and the face ratio was 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 their magnitude could be compared to 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 3-4). Across the studies, judgements of aggression shared significantly stronger associations with the face ratio than judgements of trustworthiness. A one sample t-test was also used to test whether the differences in correlation magnitude across studies was

[11] 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 to 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 1000 ms after which the aforementioned question appeared), it may provide some insight as to whether the phrasing of the question influences judgements of aggression. When we compare the responses from participants in Study 4 (n = 40) to those from participants in our ongoing study (current n = 40), correlations between the face ratio and mean judgements of aggression were not significantly different (Study 4: $r = .48$, $p < .001$; ongoing study $r = .42$, $p < .001$; Fisher r-to-z transformation to test difference between correlations: $z = 0.32$, $p = .75$) and the correlation between judgements 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 (mean = 3.56, $SD = 0.87$) did not significantly differ ($t_{128} = -0.81$, $p = .42$) from those made by participants in the ongoing study (mean = 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 judgements of aggression and the relationship these judgements of aggression share with the face ratio.

significantly different from zero. This test again revealed that judgements of aggression shared significantly stronger associations with the face ratio across studies than did judgements of trustworthiness[12,13].

Discussion

Previous studies have identified links between judgements of aggression and trustworthiness (e.g., Carré et al., 2009; Oosterhof & Todorov, 2008), as well as links between each of these judgements 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 judgement 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 judgements of aggression and trustworthiness. Based on the hypothesis that snap judgements of aggression may be more relevant for survival than snap judgements of trustworthiness, and that trustworthiness judgements may simply "piggyback" on judgements of aggression when exposure to a face is limited or reputational information is lacking, we predicted that (1) judgements of aggression would share stronger links with the face ratio

[12] Also, a repeated measures ANOVA with correlation type (aggression and face ratio vs trust and face ratio) as a within-subject factor and study number as a between subject factor revealed a main effect of correlation type ($F_{1,156} = 27.32$, $p < .001$) and of study ($F_{4,156} = 3.49$, $p = .009$), but no interaction between these factors ($p = .55$). Therefore, the association between the face ratio and judgements of aggression ($M = .32$, $SE = .02$) is stronger than the association between the face ratio and judgements of trustworthiness ($M = .22$, $SE = .02$) when controlling for any variation across study designs (e.g., stimuli, rating instructions).

[13] In Supplementary Materials, I also investigate the relationship between reaction times and the face ratio; the second time participants rate a face, regardless of the type of judgement (aggression or trustworthiness), participants are quicker to rate men with larger face ratios than men with smaller face ratios.

than judgements of trustworthiness, (2) judgements of aggression would form faster than judgements of trustworthiness, and (3) judgements of aggression would mediate the link between the face ratio and judgements 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 judgements of aggression than with judgements of trustworthiness, judgements of aggression were made more quickly than were judgements of trustworthiness, and judgements of aggression mediated the face ratio-trustworthiness link, although there was evidence of only partial mediation when multilevel modelling 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 judgements 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 behavioural dispositions important to survival (Carré et al., 2009); judgements 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 judgements of trustworthiness are relevant for social interactions and can be formed after a short exposure time to a face as well (Willis & Todorov, 2006), the encounters in which judgements 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 judgements of aggression were made more rapidly than judgements 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 judgements may be more accessible than trustworthiness judgements. Further, based on a principal components analysis, the dimensions of valence and dominance accounted for 82 percent of the variability in judgement ratings (trustworthy, emotionally stable, responsible, sociable, caring, weird, attractive, mean, intelligent, aggressive, unhappy, confident, and dominant) (Oosterhof & Todorov, 2008). Judgements of trustworthiness loaded only on the dimension of valence, whereas judgements of aggression loaded onto both dimensions. Thus, more relevant inferences about the characteristics within each dimension can be made from judgements of aggression than from judgements of trustworthiness. Further, participants reported higher confidence in judgements of aggression than in those of trustworthiness (see bottom panel of Figure 3-2 of Willis & Todorov, 2006). Thus, judgements 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 judgements of trustworthiness.

Our results for the speed with which judgements of aggression and trustworthiness are made were opposite to those reported by Willis and Todorov (2006).

Specifically, the middle panel of Figure 3-2 in their manuscript suggests that judgements of trustworthiness are made faster than are judgements 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 judgements could be made, whereas participants in Willis and Todorov (2006) could only provide a judgement after viewing photographs for durations of 100 to 1000 ms; thus, they may have inadvertently prevented the recording of genuine, rapid responses.

Our results also provide insight as to why high level cognitive judgements, 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, judgements such as trustworthiness may piggyback on, or extract information from, more primitive or survival-relevant social judgements such as those of aggression, which are better suited for encounters with strangers. Thus, as opposed to judgements of aggression and trustworthiness sharing parallel perceptual processes such that facial features lead to the simultaneous formation of these judgements, it instead appears that these judgements are formed sequentially; facial features cue judgements of aggression, which in turn influence subsequent judgements of trustworthiness. Future studies may benefit from examining whether these results are generalizable to other social judgements that differ with regards to primacy or cognitive complexity (e.g., judgements of competence/intelligence versus threat/danger/desirability).

It should be noted that although these mediation and speed of rating effects were relatively strong in the current study, judgements of aggression based on the face likely play less of a role in shaping judgements 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 judgements of trustworthiness.

Some researchers have proposed that the relationship between the face ratio and judgements of, or actual, behaviour 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 to 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, judgements of aggression were related to the face ratio when judgements of anger were controlled statistically, indicating that the face ratio guides judgements of aggressiveness over and above the extent to which it is associated with perceptions of anger. Additionally, although observers' judgements 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 behaviour among men with larger face ratios compared to 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 behaviour 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 signalling mechanism common to both human and non-human primates. A recent study found that capuchins with larger face ratios exhibited more dominant behaviour (Lefevre et al., 2013) and assertiveness (Wilson et al., 2014) than those with smaller face ratios. Nevertheless, whether non-human primates use this signal has yet to be investigated. We have found that judgements of aggression made by both children and adults are associated with the face ratio, even when the judgements 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, judgements 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 through 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 judgements 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 behaviours (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 behaviour (e.g., Farrell & McGinnis, 2004). There is also some evidence in humans that testosterone concentrations measured during early adolescence are predictive of antisocial behaviour 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 behavioural 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 judgements (aggression and trust) than were those assigned to the condition in which trustworthiness was rated first. This group difference in response time was likely caused by these initial aggression 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 judgements. 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, judgements of "aggression" facilitated subsequent judgements, whereas judgements of "trustworthiness" impeded subsequent judgements, perhaps because aggression requires speed whereas trust requires deliberation.

It should also be noted that irrespective of the ultimate mechanisms (e.g., rapid judgements of aggression may be more relevant for survival than rapid judgements 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 judgements; 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 “aggressive” and “mean” than as “trustworthy. In addition, Google’s Ngram Viewer (as used in Greenfield, 2013; Michel et al., 2011) reported the frequency of either the single word “aggressive” or the phrases “aggressive looking” or “aggressive face” since 1985 to be 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 reaction time (reviewed in Borowsky & Besner, 1993).

Limitation

One limitation to these studies is that we did not include measures of the men’s actual aggressive or untrustworthy behaviour. Thus, we cannot determine whether or not judgements of aggression or of trustworthiness provided in the current study were accurate. In any case, despite variability in the extent to which social judgements are accurate (see Rule, Krendl, Ivcevic, & Ambady, 2012; Olivola & Todorov, 2010; Zebrowitz & Montepare, 2008), they nonetheless modulate behaviour 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 judgements are important aspects of human psychology, independent of their accuracy. There are many examples in which facial features facilitate social judgements 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 focussing on possible errors in human judgements, it is worthwhile to examine the cognitive and perceptual systems that produce these judgements, as they likely served an adaptive purpose in the past (Haselton & Funder, 2006), thus leading to their maintenance and consistency in promoting social judgements. The association between the face ratio and judgements 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 behaviour[14], it influences the observer's behaviour (e.g., likelihood of sharing or not sharing a resource with the individual, Haselhuhn et al., 2013; likelihood of trusting or not trusting the individual in an economic interaction, Stirrat & Perrett, 2010) and thus is an important psychological/perceptual process worthy of investigation.

Conclusion

Judgements of trustworthiness are critical in regulating our social interactions and have been shown to modulate behaviour 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 judgements of trustworthiness. We extend these findings, however, and show that this link is not direct, but is instead mediated by judgements of aggression. Consistent with this mediation model, temporal precedence was also established: Judgements of aggression were provided faster by participants than were judgements of trustworthiness. Therefore,

[14] Which, some evidence suggests, it does (e.g., Carré & McCormick, 2008; Goetz et al., 2013; Stirrat & Perrett, 2010).

instead of judgements of aggression and trustworthiness forming simultaneously after the perception and processing of a face, it is more likely that judgements of aggression form first, prompting the subsequent formation of judgements of trustworthiness. As such, the formation of many social judgements, even those that are highly correlated, may be best characterized by a sequential "piggyback" framework wherein cognitively complex judgements occur after, and extract information from, more primitive social judgements, especially when exposure to facial information is brief and reputational information is absent.

Tables

Table 3-1.

Descriptive statistics for the face ratio and for judgements provided by observers.

Measure	Means (*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

Table 3-2.

Sobel tests of mediation across studies. Significant values indicate that aggression reduced the association between the face ratio and trustworthiness.

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

Figures

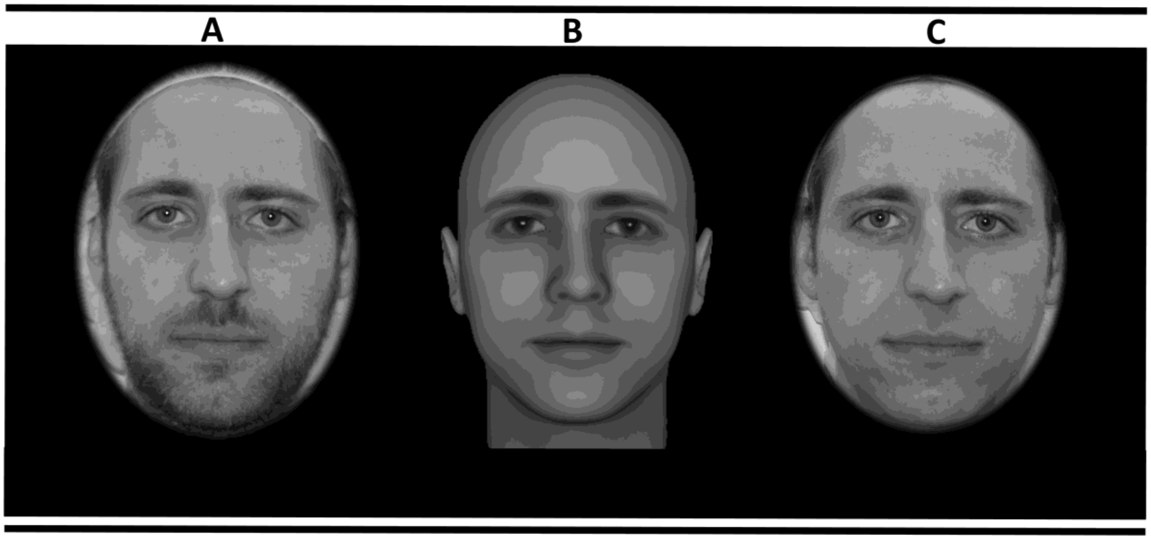

Figure 3-1. Panel A shows an example of a male face with facial hair. Panel B shows the same male after it was loaded into Facegen and the textural details were removed. Panel C shows the same male face after the man shaved, to allow for comparison to Panel B. The figure also shows how real faces and Facegen faces were presented to observers.

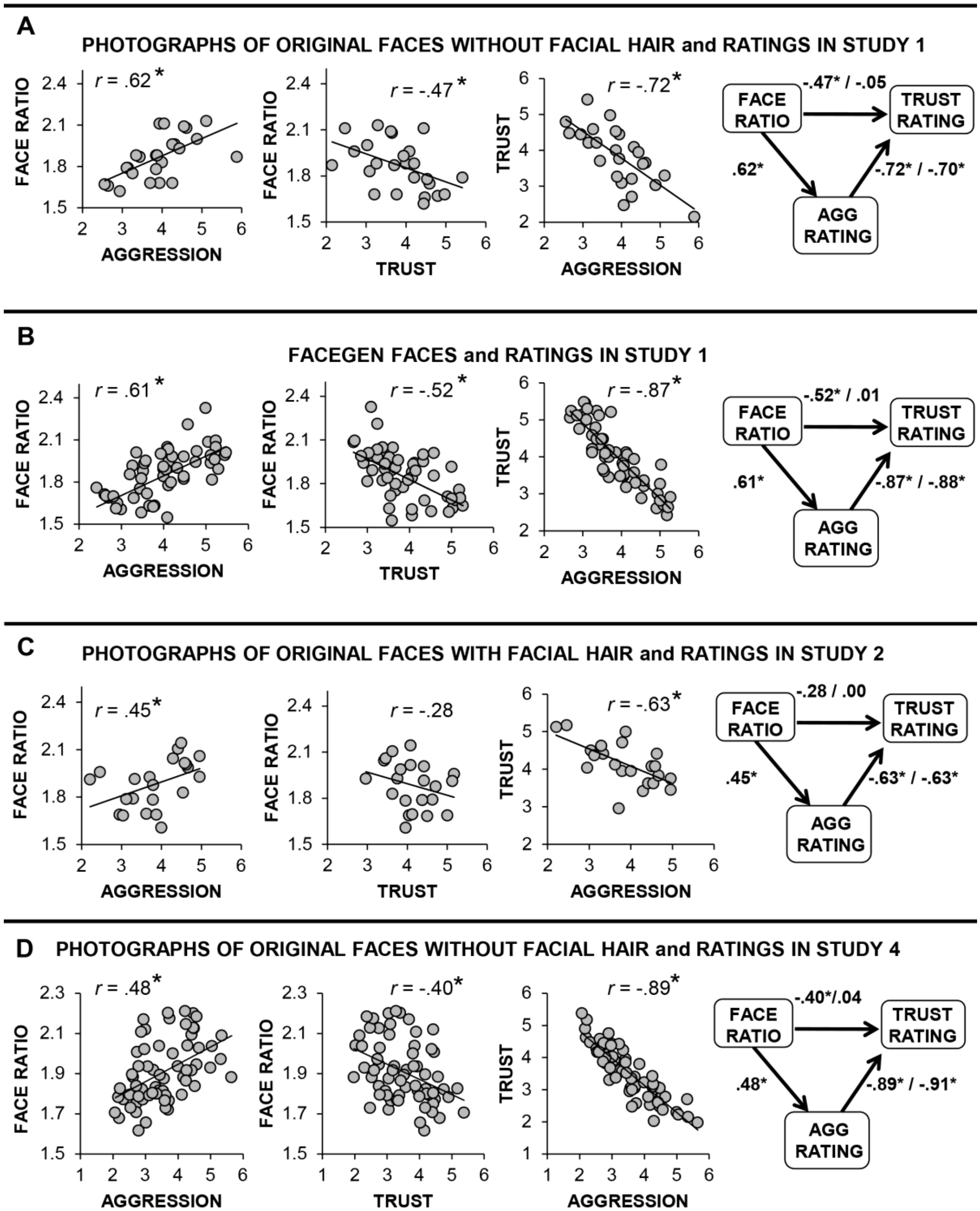

Figure 3-2. Scatterplots of the face ratios and observers' mean judgements of aggression and trustworthiness are displayed in the three left-most figures of panels A, B, C, and D. The right-most figure in each panel shows the results of mediation analyses used to

determine if the relationships between the face ratio and judgements of trustworthiness were mediated by judgements of aggression (AGG). The numbers shown are standardized regression coefficients (β weights). The first β weights between the face ratio and judgements of trustworthiness and between judgements of aggression and judgements 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 judgements of aggression were entered as simultaneous predictors of judgements of trustworthiness.

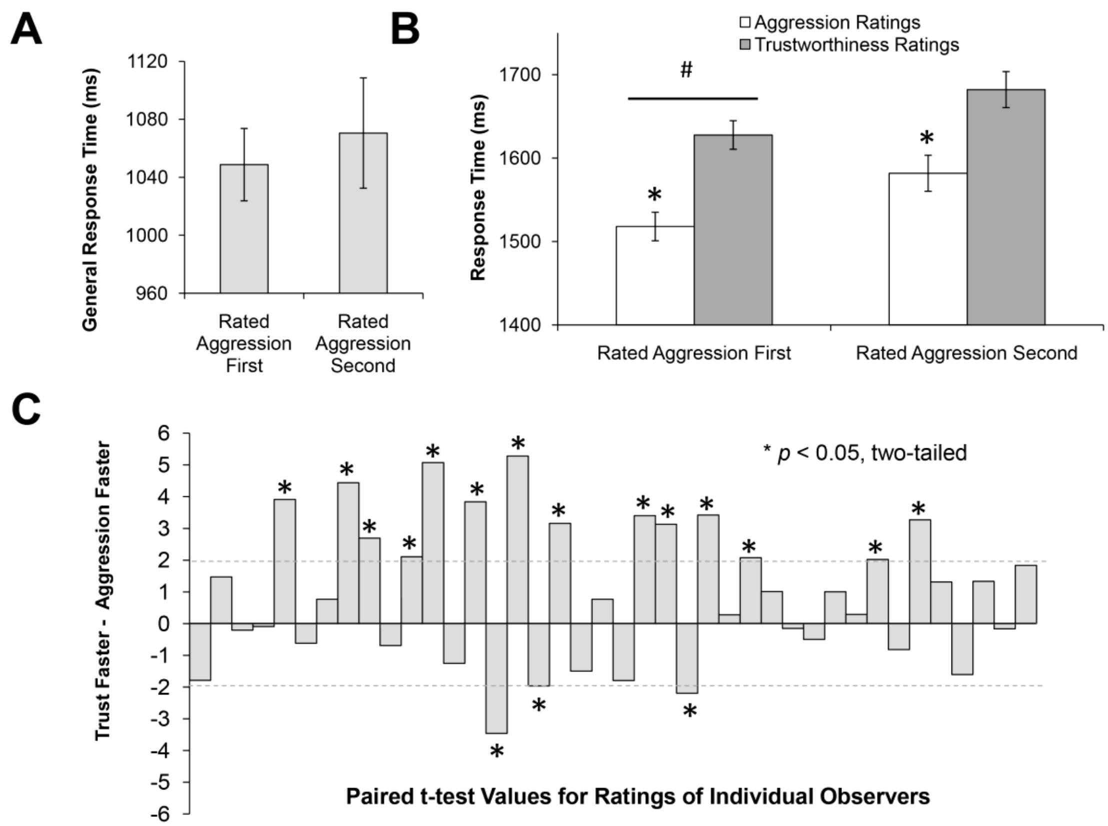

Figure 3-3. Panel A shows the participants' mean response speed in Study 4 when asked to rate the size of circles as quickly as possible. The bar on the left represents the mean and S.E.M. of participants who were asked to judge aggression first (and trustworthiness second) whereas the bar on the right represents the mean and S.E.M. of participants who were asked to judge aggression second (and trustworthiness first). Panel B shows participants' mean speed of judgements of aggression (white bars) versus judgements of trustworthiness (grey bars) as a function of whether they rated aggression first or second. * = Aggression judged faster than trustworthiness, $p < 0.05$. # = Judgements were faster if participants rated aggression first rather than second, $p < 0.05$. Panel C shows paired t-test values comparing the speed of aggression versus trustworthiness judgements within each observer. Positive values indicate that judgements of aggression were quicker

whereas negative values indicate judgements of trust were quicker. Dashed lines represent critical t-values ($p < 0.05$, two-tailed).

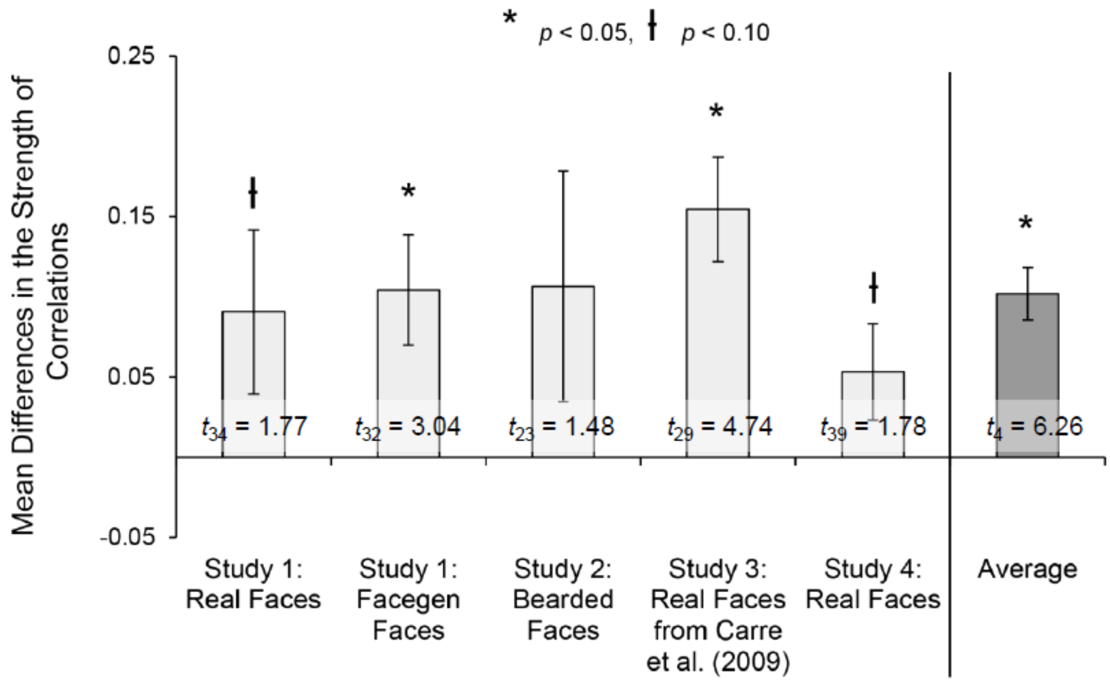

Figure 3-4. Bar graph showing differences in the strength of the associations between the face ratio and judgements of aggression compared to the strength of the associations between the face ratio and judgements of trustworthiness. Bars represent the mean difference in correlation strength such that values above the horizontal axis indicate stronger associations between the face ratio and judgements of aggression than between the face ratio and judgements of trustworthiness. The lighter bars show the mean difference within each study (and the corresponding paired-sample *t*-test values comparing the strength of the correlations) whereas the dark bar shows the average difference across studies (and the corresponding one-sample *t*-test value comparing the difference score across the studies, compared to a value of zero). Error bars represent S.E.M.s.

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. *British Journal of Psychology*. doi: 10.1111/bjop.12035

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

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 & 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). ACM Press/Addison-Wesley Publishing Co.

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. 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 of the Royal Society B: Biological Sciences, 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/p6543

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: 10.117/0956797613504965.

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. *Biodemography and 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*. 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. (2012). 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: Macmillan.

Haselhuhn, M. P. & Wong, E. M. (2012). Bad to the bone: Facial structure predicts unethical behaviour. *Proceedings of the Royal Society B: Biological Sciences, 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*, 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*, *14,* 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*. doi: 10.1002/per.1942

Lefevre, C. E., Wilson, V. A. D., Morton, F. B., Brosnan, S., Paukner, A., & Bates, T. C. (2013). Facial width-to-height ratio predicts dominance and status in the primate face. Poster presented at the 8th Annual Conference of the European Human

Behaviour and Evolution Association, Amsterdam, NL. Poster retrieved from http://www.ehbea2013.com/ ehbea_details/upload/list_posters.php

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: 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. *The 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.

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 and 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. *Proceedings of the National Academy of Sciences*, *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. 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. (2012). Accuracy and consensus in judgments of trustworthiness from faces: Behavioral and neural correlates. *Journal of Personality and Social Psychology, 104,* 409-226. doi: 10.1037/a0031050

Ruppert, D. (1988). Trimming and Winsorization. *Encyclopedia of statistical sciences*. 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 of the Royal Society B: Biological Sciences*, *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]. 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.org/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

Trebický, V., Havlícek, 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. *Encyclopedia of statistics in behavioral science*. 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, 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

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

Wilson, V.A.D., 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

Rationale for Chapter 4

In Chapter 3, I provided evidence that the face ratio is more strongly and directly linked to judgements of aggression than to judgements of trustworthiness. This finding may provide clues as to the primary social function of the face ratio in interpersonal interactions. For example, rather than observer sensitivity to this metric emerging because of its importance in regulating trust decisions in economic and cooperative interactions (social decisions which may instead depend on reputation or long-term, repetitive interactions), these results suggest that sensitivity to the face ratio may have instead evolved because such sensitivity facilitated judicious fight, flight, and submit decisions in competitive and agonistic contests. Although not investigated in Chapter 3, it is also possible that both of these judgements tap into a broader threat potential construct, of which aggression may simply be a better indicator than trust.

In Chapter 4, I further investigate the use of this metric when observers assess threat potential. One limitation of most studies on social perception to date is the use of clean shaven rather than bearded men as stimuli. If the face ratio and sensitivity to it are indeed part of an evolved advertisement and assessment system of threat potential, then one would expect observers to utilize the metric when judging the threat potential of men who are bearded. If there is no association between judgements of aggression and the face ratio of bearded men, then it is unlikely that the sensitivity to the face ratio is part of an evolved mechanism; many interactions today and likely in our evolutionary past involved bearded men. Therefore, based on the hypothesis that the face ratio and sensitivity to it is part of an evolved mechanism designed for advertising and assessing threat, I predicted that the face ratio would be associated with judgements of aggression in bearded men.

Although I already provided an initial test of this prediction in one of the studies in Chapter 3, I provide a more rigorous test in Chapter 4, using bearded and non-bearded versions of the same male faces.

Chapter 4: Facing our ancestors: Judgements of aggression are consistent and related to the facial width-to-height ratio in men irrespective of beards

The published version of this chapter is:

Geniole, S.N. & McCormick, C.M. (2015). Facing our ancestors: Judgments of aggression are consistent and related to the facial width-to-height ratio in men irrespective of beards. *Evolution and Human Behavior, 36*(4), 279-285. doi: 10.1016/j.evolhumbehav.2014.12.005

Author Contribution: I was the primary investigator of this work, responsible for the majority of the study design, data analysis, and writing of the manuscript.

Introduction

The ability to accurately assess dominance, strength, and aggressive potential has high adaptive value; across a variety of species, rivals can assess each other's formidability, which informs decisions to defer or contend in competitive interactions (reviewed in Blanchard et al., 2011; van Staaden et al., 2011). Sensitivity to threat does not appear to depend on experience; rhesus monkeys (*Macaca mulatt*a) reared in isolation respond to facial displays of threat (Sackett, 1966). There is evidence that the human perceptual system also is tuned to cues of threat in the face. For example, faces displaying angry expressions are detected more quickly than are faces displaying happy expressions (Hansen et al., 1988; Ohman et al., 2001). Judgements of threat and aggressiveness are made more rapidly than are judgements of intelligence (Bar et al., 2006) and trust (Geniole, Molnar, et al., 2014), which suggests traits related to survival form faster than do others. There is high consistency across observers in judgements of dominance, strength, and aggressive potential from views of men's faces, and such judgements have accuracy (e.g., Carré et al., 2009; Sell et al., 2009). For example, observers were able to accurately gauge the upper body strength (Sell et al., 2009), aggressiveness (Třebický et al., 2013), and toughness (Zilioli et al., 2014) of men, and judgements of dominance were correlated with the actual strength of men (Undurraga et al., 2010), simply from viewing photographs of their faces. Further, humans accurately judge dominance in chimpanzee (*Pan troglodytes*) faces, which suggests that the features responsible for cueing such accurate judgements may be common to both species (Kramer et al., 2011; Kramer & Ward, 2012). Some of the cues that are associated with judgements of dominance and strength in humans are size of the chin (Burton & Rule, 2013) and jaw (Keating et al., 1981; Berry & McArthur, 1986), cues that also are associated with judgements of masculinity. Nevertheless, such cues in the lower face would be masked by beards, which suggests there may be features in the face that cue such judgements irrespective of facial hair.

One facial metric relevant for assessments of threat that may not be masked by beards is the facial-width-to height ratio (FWHR; the bizygomatic width divided by upper face height, the distance between brow and upper lip; Weston et al., 2007). There is evidence that the FWHR is an accurate indicator of aggression [Carré & McCormick, 2008; Goetz et al., 2013; Lefevre et al., 2014; Třebický et al., 2014; Welker et al., 2014; Zilioli et al., 2014; although some authors have reported marginal (*p* = .057, Deaner et al., 2012) or non-significant associations (Gómez-Valdés et al., 2013; Özener, 2011)]. Observers' judgements of aggressiveness and dominance (which are highly correlated, *r* = .92, Carré et al., 2009) in neutrally posed faces are highly and consistently correlated positively with the FWHR (Alrajih & Ward, 2014; Boshyan et al., 2013; Burton & Rule, 2013; Carré et al., 2009, 2010; Geniole et al., 2012; Geniole, Molnar, et al., 2014; Lefevre & Lewis, 2013; Valentine et al., 2014), even when made by children of faces of an unfamiliar ethnicity (Short et al., 2012). Further, the face ratio predicted assertiveness (V. Wilson et al., 2014) and the achievement of alpha status (Lefevre et al., 2014) in capuchin monkeys (*Sapajus spp.*), again consistent with the possibility that similar features are used to cue accurate judgements among primates.

Based on the hypothesis that sensitivity to the FWHR is part of an evolved perceptual mechanism designed to detect aggressive potential in others, and that interactions in ancestral history likely involved bearded men[15], we predicted: (1) measurement of the FWHR would not be disrupted by facial hair, and; (2) judgements of aggression would be related to the FWHR in bearded and in non-bearded versions of men's faces. We previously reported an association between the FWHR and observers' ratings of aggressiveness in a set of men with facial hair (Geniole, Molnar, et al., 2014), but we did not have shaved versions of the same male faces for comparison. Additionally, the facial hair of the men was often highly stylized because the facial hair was grown for participation in "Movember" (to support prostate cancer research). We addressed these limitations here by using photographs taken both before and after a prolonged period of beard growth.

The experiment also provided the opportunity to test some additional predictions. We predicted that beardedness may drive judgements of masculinity to a greater extent than the face ratio based on evidence of positive associations between extent of facial hair and masculinity (Dixson & Brooks, 2013). As a control judgement, we also had participants rate the attractiveness of the faces, given that such judgements tend to share

[15] The extent to which early ancestors were bearded is unknown, but given common ancestry with other primates, at some point our evolution as a species involved pressure to reduce the coarseness and visibility of bodily hair on many parts of the body (Montagna, 1985) while accentuating growth and maintaining visibility of head hair and, in men, some facial hair (Darwin, 1874/2002). Evidence depends largely on the archaeological record of tools and art and early writings. The earliest depiction of a male figure is considered to be the Lion Man of Hohlenstein (~38,000 BCE) (Knight et al., 1998), named for the lion-like rather than human face, with less stylized depictions appearing rather recently in an evolutionary scale. For example, neolithic (~8000 BCE) art from Nehal Hemar has depictions of faces with beards (Mithen, 2004), there is evidence of tools for shaving in the late predynastic period of Egygt (pre 3000 BCE) (Scheel, 1989), Akkadian art from Mesopotamia depicts men with long, stylized beards (http://www.metmuseum.org/toah/works-of-art/L.1992.23.5), the Torah (~1300 BCE) has prohibitions against shaving (Olyan, 1988), the *Shatapatha Brahmana* (~700 BCE) from India makes reference to the shaving of facial hair (Hiltebeitel & Miller, 1998), and several of the Terracotta Warriors (~200 BCE) from China wear beards (Deng, 2011). Nevertheless, the ability to grow a beard varies across ethnicities (e.g., Darwin, 1874/2009), time of year (Randall & Ebling, 1991), and the lifespan (Hamilton, 1958). We propose that if the FWHR is part of an evolved system designed for communicating threat, this feature should be perceivable by observers and influence their social judgements irrespective of facial hair. If the current visual system cannot readily extract this cue in the presence of facial hair then it is unlikely that the face ratio influences social judgements across all human populations and periods.

weak and inconsistent links with the FWHR (e.g., Carré et al., 2009; Geniole & McCormick, 2013; Geniole et al., 2012; Valentine et al., 2014). Beardedness, however, is known to influence judgements of attractiveness (e.g., Dixson & Brooks, 2013; Muscarella & Cunningham, 1996; Dixson & Vasey, 2012; Neave & Shields, 2008). Thus, we expected that attractiveness would not be related to the FWHR in either bearded or non-bearded versions of men's faces, but would be related to the amount of facial hair.

Methods

Participants

Participants (34 women, 22 men; *M* age = 19.89 years, *SD* = 3.44; 73% White, 4% Black, 7% Asian, 16% other) were undergraduates recruited through an online research pool at Brock University and received a course credit for participation. We aimed for 60 participants [*n* = 30 per order of ratings, see section (d)]. Of the 60, four participants are not included here: three participants because of equipment failures during the experiment that compromised data collection and one because he used the same rating for each face (face rating for every face was rated as "1"). All participants consented to the procedures of the study, which were approved by the Brock University Research Ethics Board.

Stimuli

The stimuli set consisted of 25 faces photographed with and without a beard, retrieved from www.youtube.com (see Supplementary Table 4-1 for links to each video). The faces were found by searching the website for videos of men who documented their beard growth by taking photographs of their face at various time points throughout a prolonged period of sustained beard growth. A research assistant blinded to the hypotheses searched for the videos using combinations of keywords such as “beard”, “facial hair”, and “picture a day”. Screenshots of the videos were taken at two time points, before and after substantial beard growth. Only screenshots from videos of men who appeared Caucasian, younger than 40 years of age (to avoid any ethnicity- and age-based stereotypes in judgements), were facing the camera directly, and were posed in neutral expressions for the two screen shots were included. With these inclusion criteria, the stimuli set consisted of 50 facial photographs of 25 faces, each face photographed twice, once with a beard and once without. The stimuli were standardized as in Carré et al. (2009).

Measurement of the FWHR

A research assistant blind to the hypotheses of the study measured and calculated the FWHRs of the stimuli using the procedures described in Carré and McCormick (2008). Specifically, a measure of the width of the face, from zygion to zygion, was divided by a measure of the height of the upper-face, the distance from the mid-brow to the upper lip. To avoid biases in measurement of the bearded faces (and to avoid artificially exaggerating the association between the face ratio measured in non-bearded and bearded faces), the research assistant was shown the measurement points of the face ratio in a photograph of a non-bearded face. Afterwards, the research assistant was asked to simply estimate the other points for the face ratio in all the other faces (both bearded and non-bearded). In all but one case, the upper lips of the men were visible and thus it is unlikely that the beards of the men obscured these measurement points.

Judgement of Faces

Faces were presented and rated using E-Prime software and laptop computers. Participants rated all of the faces within a given face version (bearded versions vs. non-bearded versions) once for each of the three blocks (aggression, masculinity, then attractiveness). Because the main research question was about the relationship between judgements of aggression and the FWHR, all participants made judgements in the same order. Nevertheless, we previously found no effect of order on sets of ratings of aggressiveness, trustworthiness, masculinity, and/or attractiveness (Geniole & McCormick, 2013; Geniole, Molnar, et al., 2014). The order, however, in which participants rated the two face versions (bearded or non-bearded) was counterbalanced across participants. After each face within a version was rated three times, once for each rating, the next version was rated using the same procedure.

Within each block, participants were asked to provide the ratings for each face as quickly as possible, using their gut instincts, and keys 1 through 7 on a laptop keyboard. The specific questions, which appeared at the beginning of each corresponding block (aggression, masculinity, and attractiveness block) were: “How aggressive does this person look?” (1 = *not at all aggressive*, 7 = *very aggressive*); “How masculine does this person look?” (1 = *not at all masculine*, 7 = *very masculine*), and; “How attractive does this person look?”(1 = *not at all attractive*, 7 = *very attractive*). As a manipulation check, after participants rated both versions of the faces on the three characteristics, the faces from each version were combined, randomized across participants, and rated for hairiness. The specific question was “How much facial hair does this person have?” (1 = *none at all*, 7 = *very much*). A white fixation cross was displayed for 500 ms before the presentation of each face. Once a face appeared, it remained on the screen until the participant provided a response, at which point the fixation cross would reappear for 500 ms, and then the next face within the given stimuli set would be displayed. The order of the presentation of faces was randomized within each block across participants.

Therefore, every participant rated both the bearded and non-bearded versions of the faces, and the order of ratings (bearded versions first or non-bearded versions first) was counter-balanced. Specifically, after rating each face of a given version on aggression, masculinity, and attractiveness, the participants rated the second version of the faces on the same characteristics (the order of the faces was randomized across participants for both versions but the order of the ratings – aggression, masculinity, and attractiveness – was kept constant). After rating both versions of the faces on all three characteristics, all participants rated a fourth characteristic, hairiness of the faces. For this rating, both versions of the faces were combined and the order of the faces was randomized across participants. Therefore, every face (bearded and non-bearded) was rated four times by every participant: once for aggression, once for masculinity, once for attractiveness, and once for hairiness.

Statistical Analyses

Analyses consisted of paired-samples t-tests and Pearson product moment correlations. When comparing the magnitude of the correlations, values were tested using Steiger's Z or, for individuals, were transformed to Fisher's Z values and analysed using one-sample or paired-samples t-tests, or repeated measures ANOVAs. Alpha value of $p <$ 0.05, two-tailed, was used to determine statistical significance. Analyses were conducted with the sexes combined because judgements made by men were highly correlated with those made by women (e.g., men's judgements of aggression in bearded faces were highly correlated with women's judgements of aggression in the same bearded faces; the same was true for every other judgement type; all *r*s > .86, all *p*s < 0.001).

Results

Manipulation check

The bearded versions of the men's faces were rated as hairier than the non-bearded version (t_{24} = 18.35, $p < 0.001$, *Cohen's d* = 3.68, *M* difference = 3.68, *95% CI* difference = 3.27 to 4.10; see Table 4-1). There was no overlap in the two distributions.

The FWHR can be measured reliably when faces have facial hair

The FWHR of the men did not differ between bearded and non-bearded versions (t_{24} = 1.40, p = 0.17, see Table 4-1)[16], and the measurements of the two versions were highly correlated (r = .93, $p < 0.001$). Therefore, for subsequent analyses involving the FWHR, we used an average of the bearded and non-bearded FWHR measurements (M = 1.67, SD = 0.15)[17].

Observers' judgements of aggression are associated with the FWHR in both bearded and non-bearded versions of men's faces

[16] The FWHR in this sample of faces was smaller than we have found previously in larger data sets (Skorska et al., in press). There is the possibility that there is some distortion in the displaying of YouTube videos, as the variability in this data set is similar to that obtained previously.

[17] Because sideburns varied across the photos and may have been bigger among the bearded versions of the faces, we asked another research assistant to calculate the face ratios using a measurement of width that ended at the extremity of the zygion but excluded the sideburns, and to calculate the face ratios again using a measurement of width that ended at the extremity of the sideburn, thus including both the zygion and the sideburns. Both versions (sideburns vs no sideburns) of the face ratio measure were highly correlated (r = .97. $p < 0.001$); as expected, the one including the sideburns was larger (t = 8.79, $p <$ 0.001). Both versions were highly correlated with the face ratio measure provided by the first research assistant (with sideburns: r = .97; without sideburns: r = .99). When both versions (sideburns vs no sideburns) were entered as simultaneous predictors of the first research assistant's FWHR measure, the version without sideburns was significant ($p < 0.001$) whereas the version with sideburns was not (p = .48) and accounted for little of the variability above and beyond that of the version excluding sideburns (sr = .02). These results indicate that the sideburns did not significantly obscure or bias the measurement of the face ratios performed by the first research assistant. Additionally, the high correlation between the measures of the face ratio made by the first research assistant, as well as the two made by the second research assistant, speak to the reliability with which this metric can be perceived in faces with facial hair.

Average judgements of aggressiveness in the bearded and non-bearded versions were correlated (*r* = .78, $p < 0.001$; see Figure 4-1-A), but the bearded versions were judged as significantly more aggressive ($t_{24} = 5.12$, $p < 0.001$, *Cohen's d* = 1.03, *M* difference = 0.65, *95% CI* difference = 0.39 - 0.91) than the non-bearded versions (see Table 4-1). Judgements of aggression were associated with the FWHR in both non-bearded (*r* = .66, $p < 0.001$) and bearded faces (*r* = .59, *p* = 0.002), and the strength of these two correlations did not significantly differ (*Steiger's Z* = -0.66, *p* = 0.75) (see Figure 4-1-B).

Observers' judgements of masculinity, and the relationship between masculinity and the FWHR, are disrupted by beards

There was no association between average judgements of masculinity made for non-bearded and for bearded versions (*r* = .18, *p* = 0.40; see Figure 4-1-C), and the higher ratings of masculinity for bearded than non-bearded versions did not meet statistical significance ($t_{24} = 1.91$, *p* = 0.07, *Cohen's d* = 0.39, *M* difference = 0.38, *95% CI* difference = -0.03 to 0.78; see Table 4-1). Judgements of masculinity were associated with the FWHR in the non-bearded faces (*r* = .55, *p* = 0.004), but this association was weaker (*Steiger's Z* = 1.89, *p* = 0.03) and non-significant in the bearded faces (*r* = .08, *p* = 0.71) (see Figure 4-1-D).

Observers' judgements of attractiveness are consistent in both bearded and non-bearded versions of men's faces, but not associated with the FWHR

Average judgements of attractiveness in the bearded and non-bearded versions were correlated (*r* = .64, *p* = 0.001; see Figure 4-1-E), and the lower ratings of attractiveness for bearded than for non-bearded versions did not meet statistical significance (t_{24} = 1.95, p = 0.06, *Cohen's d* = 0.39, *M* difference = -0.24, *95% CI* difference = -0.50 to 0.01; see Table 4-1). Judgements of attractiveness were not associated with the FWHR in either version (bearded: r = .25, p = 0.23; non-bearded: r = .33, p = 0.11) (*Steiger's Z* = 0.47, p = 0.68) (see Figure 4-1-F).

Bivariate correlations between all of the study variables are reported in Table 4-2[18].

Supplementary analyses

We conducted additional analyses to investigate the extent to which the relationships observed are evident at the level of individual observers and not only in averaged ratings (McCormick, 2013) and results at the level of the individual were consistent with the analyses reported here using the ratings averaged across observers (see supplementary figures and text). The relationship between judgements of attractiveness and judgements of hairiness are also explored in the supplementary analyses.

Discussion

If sensitivity to the FWHR is part of an evolved mechanism to detect aggressive potential, and interactions in ancestral history involved men with facial hair, judgements of aggression should be associated with this metric when men's faces are bearded. Observers' judgements of aggression of bearded faces were highly correlated with those

[18] I also investigated, in both the bearded and non-bearded versions of the men's faces, the extent to which the face ratio predicted judgements of aggression controlling for judgements of masculinity, attractiveness, and hairiness. The face ratio was a significant predictor in both versions, albeit less strong in the non-bearded versions (bearded: β = .62, semi-partial r = .57, p = .001; non-bearded: β = .39, semi-partial r = .31, p = .001), likely because lower-face cues related to masculinity and dominance are more visible in the non-bearded than the bearded faces, thus playing a greater role in influencing judgements of aggression in these non-bearded versions of the faces. Indeed, judgements of masculinity were much more strongly related to judgements of aggression in the bearded than in the non-bearded versions of the men's faces.

made of non-bearded faces, suggesting the use of a similar perceptual strategy for both versions (i.e., the use of similar cues for assessing aggression in both versions of faces). There was no difference in the FWHR measured in the two versions, indicating that beard growth does not obscure the FWHR. Additionally, the judgements of aggression were positively correlated with the FWHR in both face versions, and the strength of these two correlations did not differ. Thus, the FWHR is a perceptual cue used to judge aggressive potential that is not disrupted by the presence of facial hair, supporting the hypothesis that sensitivity to this feature for assessment of threat would have been possible for our human ancestors even when bearded.

Beards are proposed to have evolved through sexual selection as a cue of threat, perhaps to increase the apparent size of the jaw (Guthrie, 1970), or as a marker of the association between testosterone and aggression in men (Muscarella & Cunningham, 1996; Neave & Shields, 2008). Beards are cues of masculinity and a sign of sexual maturity; beard growth begins at puberty when testosterone and its conversion to dihydrotestosterone increase (reviewed in Randall, 2008). Although there are relationships between circulating androgens and degree of beard growth (Farthing et al., 1982), whether a relationship exists between the ability to grow a beard and the FWHR could not be determined in the present study because length of time involved in the beard growth of the men was unknown. There is a report of a positive relationship between circulating testosterone and the FWHR (Lefevre et al., 2013). Thus, whether beards serve to enhance static cues in the face or to mask static cues in the lower face (e.g., “make a weak face look stronger”, Morris, 2012) is unknown, but such evidence would speak to its functional purpose.

Whereas it is unknown whether judgements formed on the basis of facial hair have accuracy (Dixson & Vasey, 2012), there is evidence that the FWHR may be an honest cue of aggression (e.g., Carré & McCormick, 2008; Goetz et al., 2013; Lefevre et al., 2014; Třebický et al., 2014; Welker et al., 2014; Zilioli et al., 2014) and other traits related to antisocial tendencies and dominance (e.g., Geniole, Keyes, et al., 2014; Haselhuhn & Wong, 2012; Mileva et al., 2014; Stirrat & Perrett, 2010) in men. Cultural and historical context are significant factors in social judgements made of bearded men. Nevertheless, across many studies, beards enhance ratings of dominance, masculinity, and aggression (reviewed in Dixson & Vasey, 2012). The extent to which beards are associated with judgements of attractiveness, however, are more variable (reviewed in Neave & Shields, 2008). In the present study, although judgements of attractiveness were lower for bearded than for non-bearded versions of the faces, the judgements for the two versions were highly correlated (see supplementary materials for more discussion of the results for attractiveness judgements in the present study).

The same sensitivity to the FWHR for assessments of aggression and attractiveness in bearded as in non-bearded men may be because of a perceptual bias to extract information preferentially from the upper-face. Such a bias would be adaptive as it would be less susceptible to masking effects of facial hair. Indeed, facial kin recognition (e.g., Dal Martello & Maloney, 2006) and the perception of emotions related to threat (e.g., Bassili, 1979) have greater accuracy when the upper-face is visible and the lower-face is masked. In contrast to judgements of aggression, the correspondence between judgements of masculinity made of bearded and non-bearded versions of men's faces was low, likely because lower-face features such as jaw and chin size contribute to perceptions of masculinity (e.g., Windhager et al., 2011). The FWHR-masculinity link was attenuated and non-significant in bearded faces, likely because this relationship is mediated through the association that the FWHR shares with lower-face features such as chin size (e.g., Burton & Rule, 2013).

Although we have proposed that the FWHR may be a signal comparable to honest signals in other species that predict aggressive intent (e.g., Morestz & Morris, 2006; Laird & Vehrencamp, 2008), it is possible that the relationship between the FWHR and trait judgements reflect emotion overgeneralization (e.g., faces with a high FWHR look angry) (Oosterhof & Todorov, 2009; Zebrowitz et al., 2010). Angry facial expressions are perceived as dominant and aggressive (e.g., Boshyan et al., 2013; Carré et al., 2009) and exaggerate cues of strength (Sell et al., 2014). The display of anger is thought to reduce the need for physical aggression by cueing fighting ability (Sell et al., 2014). In turn, facial expressions of anger serve to increase the FWHR (Carré et al., 2008; Marsh et al., 2014). In contrast to the direction of the relationship posited by the overgeneralization hypothesis, others (Marsh et al., 2005) have proposed that the facial expression of anger may have emerged or have been maintained because it simulates the informative social cue of dominance/aggression. In keeping with the latter possibility, tilted faces, which increase the FWHR, appear more intimidating than when non-tilted (Hehman et al., 2013). Further, participants spontaneously tilt their heads when asked to pose to look intimidating (Hehman et al., 2013). There is evidence that the association between FWHR and aggression judgements remains when controlling for judgements of “angry-looking” (Boshyan et al., 2013). Dynamic changes in posture and facial expressions allow bluffing; men with the lowest FWHRs displayed the bigger increases in FWHR when posing to look angry (lowered brow, raised lip; Marsh et al., 2014) or intimidating (head tilt; Hehman et al., 2013) than did men with high FWHRs. Such bluffing may be a means of preventing an aggressive encounter, and the possibility of bluffing a threat signal does not necessarily deter from the reliability of the signal (e.g., one proposal is that as spatial distance between opponents decreases the reliability of a threat signal increases; Számadó, 2008).

Although the FWHR metric accounts for a significant proportion of the variance in social judgements, statistical models incorporating numerous facial metrics have considerable success in predicting social judgements. For example, a recent study showed that linear combinations of 65 facial metrics predicted 58% of the variance in judgements of dominance, approachability, and attractiveness (Vernon et al., 2014). There was no control of emotional expressions in the faces modeled (thus the model involved non-static features), however, and it is unknown how such a model would fare when a proportion of the metrics are masked by facial hair. The stability in judgements of aggressiveness between and within observers in men's faces irrespective of facial hair as well as the strong relationship between judgements of aggression and the FWHR in both versions of the face suggests that the FWHR may be part of an evolved system of threat detection as it would have been perceptible in human ancestors with facial hair. Such a perceptual system may be common across primates; humans accurately judge dominance in chimpanzee (*Pan troglodytes*) faces (Kramer et al., 2011) and the FWHR predicts assertiveness and alpha status in capuchin monkeys (*Sapajus spp.*) (Lefevre et al., 2014).

Tables

Table 4-1.

Descriptive statistics for the face ratio and for judgements provided by observers.

Measure	Mean (*SD*)	Cronbach's α
Non-Bearded Versions of Faces		
Face Ratio	1.66 (0.14)	
Judgements of Aggression	3.45 (0.98)	.96
Judgements of Masculinity	4.49 (0.82)	.95
Judgements of Attractiveness	2.91 (0.80)	.95
Judgements of Hairiness	1.68 (0.59)	.96
Bearded Versions of Faces		
Face Ratio	1.67 (0.16)	
Judgements of Aggression	4.10 (0.89)	.96
Judgements of Masculinity	4.87 (0.71)	.94
Judgements of Attractiveness	2.67 (0.61)	.93
Judgements of Hairiness	5.36 (0.83)	.97

Table 4-2.

Associations between the face ratio and each judgment made of bearded and non-bearded versions of faces.

		Judgments of		
	Face Ratio	Aggression	Masculinity	Attractive
Non-Bearded Versions of Faces				
Judgements of Aggression	**.66**			
Judgements of Masculinity	**.55**	**.81**		
Judgements of Attractiveness	.33	-.03	.36	
Judgements of Hairiness	.21	.30	**.59**	**.53**
Bearded Versions of Faces				
Judgements of Aggression	**.59**			
Judgements of Masculinity	.08	.09		
Judgements of Attractiveness	.25	-.25	.08	
Judgements of Hairiness	.10	**.48**	.13	**-.61**

Note. Correlations in bold are significant ($p < 0.05$).

Figures

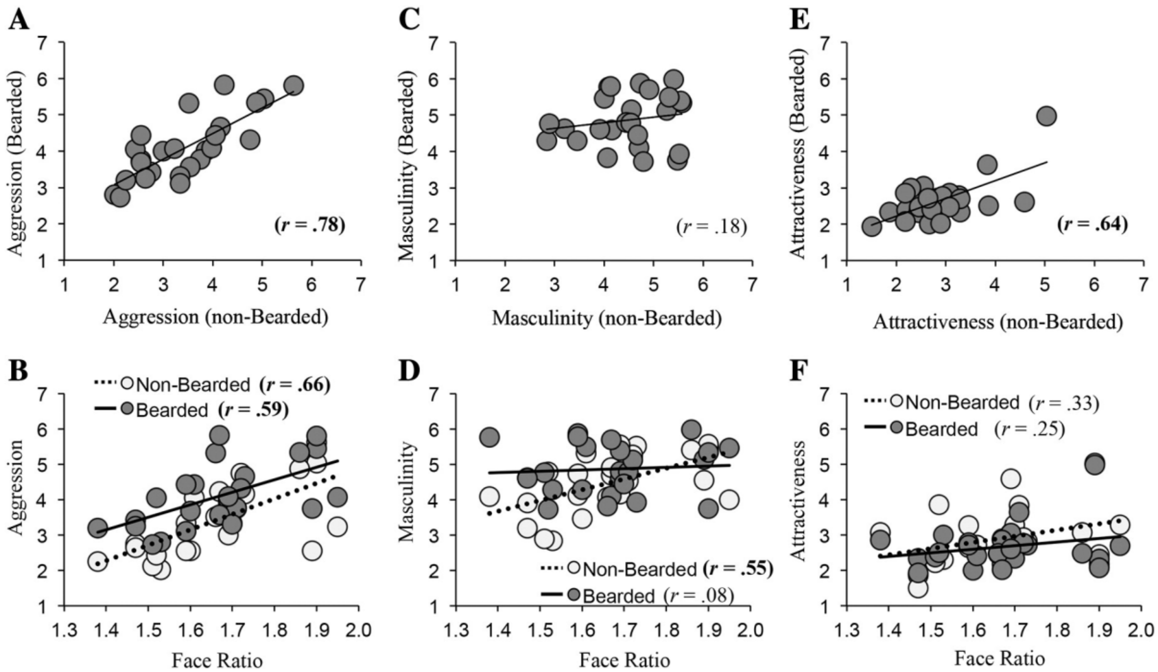

Figure 4-1. Scatterplots of the associations between: judgements of aggression made in bearded and in non-bearded versions of faces (Panel A); the FWHR and judgements of aggression made in bearded and in non-bearded versions of faces (Panel B); judgements of masculinity made in bearded and in non-bearded faces (Panel C); the FWHR and judgements of masculinity made in bearded and in non-bearded versions of faces (Panel D); judgements of attractiveness made in bearded and in non-bearded faces (Panel E), and; the FWHR and judgements of attractiveness made in bearded and in non-bearded versions of faces (Panel F). Pearson product moment correlations (*r* values) reported in boldface are significant ($p < 0.05$).

References

Alrajih, S. & Ward, J. (2014). Increased facial width-to-height ratio and perceived dominance in the faces of the UK's leading business leaders. *Brit J Psychol, 105*, 153–61. doi:10.1111/bjop.12035

Bar, M., Neta, M. & Linz, H. (2006). Very first impressions. *Emotion, 6*, 269–78. doi:10.1037/1528-3542.6.2.269

Bassili, J. N. (1979). Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. *J Pers Soc Psychol, 37*, 2049–58. doi: 10.1037/0022-3514.37.11.2049

Berry, D. S. & McArthur, L. Z. (1986). Perceiving character in faces: the impact of age-related craniofacial changes on social perception. *Psychol Bull, 100*, 3–18.

Blanchard, D. C., Griebel, G., Pobbe, R., & Blanchard, R. J. (2011). Risk assessment as an evolved threat detection and analysis process. *Neurosci Biobehav R, 35*, 991–8. doi:10.1016/j.neubiorev.2010.10.016

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. *J Gerontol B-Psychol,* 1–9. doi:10.1093/geronb/gbt054

Burton, C. M., & Rule, N. O. (2013). Judgments of height from faces are informed by dominance and facial maturity. *Soc Cog, 31*, 672–685. doi:10.1521/soco.2013.31.6.672

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. *Proc Roy B Soc, 275,* 2651–6. 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. *Psychol Sci*, *20*, 1194–8. 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/p6543

Dal Martello, M. F., & Maloney, L. T. (2006). Where are kin recognition signals in the human face? *J Vision, 6*, 1356–66. doi:10.1167/6.12.2

Darwin, C. R. (1874/2002). The descent of man, and selection in relation to sex. London: John Murray. 2d ed.; tenth thousand in John van Wyhe. (2002). The Complete Work of Charles Darwin Online, retrieved from http://darwin-online.org.uk/)

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. *J Res Pers*, *46*, 235–238. doi:10.1016/j.jrp.2012.01.005

Deng, Y. (2011). *Ancient Chinese Inventions*. Cambridge University Press.

Dixson, B. J., & Vasey, P. L. (2012). Beards augment perceptions of men's age, social status, and aggressiveness, but not attractiveness. *Behav Ecol*, *23,* 481–490. doi:10.1093/beheco/arr214

Dixson, B. J., & Brooks, R. C. (2013). The role of facial hair in women's perceptions of men's attractiveness, health, masculinity and parenting abilities. *Evol Hum Behav*, *34,* 236–241. doi: 10.1016/j.evolhumbehav.2013.02.003

Dixson, B. J. & Vasey, P. L. (2012). Beards augment perceptions of men's age, social status, and aggressiveness, but not attractiveness. *Behav Ecol*, *23*, 481–490. doi:10.1093/beheco/arr214

Farthing, M. J., Mattei, A. M., Edwards, C. R. & Dawson, A. M. (1982). Relationship between plasma testosterone and dihydrotestosterone concentrations and male facial hair growth. *Brit J Dermatol, 107*, 559–64.

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. *Pers Indiv Differ*, *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. *Evo Psych, 11*, 1027-1043.

Geniole, S. N., Molnar, D. S., Carré, J. M., & McCormick, C. M. (2014). The facial width-to-height ratio shares stronger links with judgments of aggression than with judgments of trustworthiness. *J Exp Psychol Human.* doi:10.1037/a0036732

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. *Psychol Sci*, *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

Guthrie, R. D. (1970). Evolution of human threat display organs. In *Evolutionary biology* (eds T. Dobhansky M. Hecht & W. Steers), pp. 257–302. New York, NY: Appleton-Century-Crofts.

Hamilton, J. B. (1958). Age, sex and genetic factors in the regulation of hair growth in man: A comparison of Caucasian and Japanese populations. In: Montagna W, Ellis RA, editors. The biology of hair growth. New York: Academic Press. p. 399–433.

Hansen, C. H. & Hansen, R. D. (1988) Finding the face in the crowd: An anger superiority effect. *J Pers Soc Psychol, 54,* 917–24.

Haselhuhn, M. P., & Wong, E. M. (2012). Bad to the bone: facial structure predicts unethical behaviour. *Proc Roy B Soc*, *279*, 571–576. doi:10.1098/rspb.2011.1193

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

Hiltebeitel, A. & Miller, B.D. (1998). *Hair: its power and meaning in Asian cultures*. SUNY Press.

Keating, C. F., Mazur, A., & Segall, M. H. (1981). A cross-cultural exploration of physiognomic traits of dominance and happiness. *Ethol Sociobiol*, *2*, 41–48.

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 signaling personality. *Evol Hum Behav*, *32,* 179-185. doi: 10.1016/j.evolhumbehav.2010.10.005

Knight, C., Power, C., & Mithen, S. (1998) The origins of anthropometric thinking. *J Roy Anthropolog Inst*, 4, 129-132.

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

Lefevre, C. E., Lewis, G. J., Perrett, D. I., & Penke, L. (2013). Telling facial metrics: facial width is associated with testosterone levels in men. *Evol Hum Behav, 34,* 273-279. doi:10.1016/j.evolhumbehav.2013.03.005

Lefevre, C. E., Wilson, V. A. D., Morton, F. B., Brosnan, S. F., 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

Lefevre, C. E., Etchells, P. J., Howell, E. C., Clark, A. P., & Penton-Voak, I. S. (2014). Facial width-to-height ratio predicts self-reported dominance and aggression in males and females, but a measure of masculinity does not.*Biology letters*, *10*, 20140729. doi: 10.1098/rsbl.2014.0729

Marsh, A. A., Adams, R. B., & Kleck, R. E. (2005). Why do fear and anger look the way they do? Form and social function in facial expressions. *Pers Soc Psychol Bull*, *31*, 73–86. doi:10.1177/0146167204271306

Marsh, A. A., Cardinale, E. M., Chentsova-Dutton, Y. E., Grossman, M. R., & Krumpos, K. A. (2014). Power plays: Expressive mimicry of valid agonistic cues. *Soc Psychol Pers Sci*, *5,* 684–690. doi:10.1177/1948550613519684

McCormick, C. M. (2013). Watch where and how you stick pins when playing with voodoo correlations. *J Gen Psychol, 140*, 1–5.

Mileva, V. R., Cowan, M. L., Cobey, K. D., Knowles, K. K., & Little, A. C. (2014). In the face of dominance: Self-perceived and other-perceived dominance are positively associated with facial-width-to-height ratio in men. *Pers Indiv Differ, 69*, 115–118. doi:10.1016/j.paid.2014.05.019

Mithen, S. (2004). *After the ice: a global human history, 20,000-5000 BC*. " Harvard University Press, Cambridge MA.

Montagna, W. (1985). The evolution of human skin (?). *J Hum Evol*, *14*, 3-22.

Morestz, J. A. & Morris, M. R. (2006). Phylogenetic analysis of the evolution of a signal of aggressive intent in northern swordtail fishes. *Am Nat*, *168*, 336–349. doi:10.1086/ 506920

Morris, D. (2012). *The Naked Man: A Study of the Male Body*. Vintage Books.

Muscarella, F. & Cunningham, M. R. (1996). The evolutionary significance and social perception of male pattern baldness and facial hair. *Ethol Sociobiol, 17*, 99–117.

Neave, N., & Shields, K. (2008). The effects of facial hair manipulation on female perceptions of attractiveness, masculinity, and dominance in male faces. *Personality and Individual Differences*, *45*, 373–377. doi:10.1016/j.paid.2008.05.007

Ohman, A., Lundqvist, D. & Esteves, F. (2001). The face in the crowd revisited: A threat advantage with schematic stimuli. *J Pers Soc Psychol*, *80*, 381–96.

Olyan, S. M. (1998). What do shaving rites accomplish and what do they signal in biblical ritual contexts?. *J Biblical Lit*, 611-622.

Oosterhof, N. N., & Todorov, A. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. *Emotion*, *9*, 128–33. doi:10.1037/a0014520

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

Randall, V. A. (2008). Androgens and hair growth. *Dermatol Ther*, *21*, 314–28. doi:10.1111/j.1529-8019.2008.00214.x

Randall, V. A., & Ebling, F. J. G. (1991). Seasonal changes in human hair growth. *Brit J Dermatol*, *124*, 146-151.

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

Scheel, B. (1989). *Egyptian Metalworking and Tools*. Shire Publications LTD, Riseborough UK.

Sell, A., Cosmides, L., & Tooby, J. (2014). The human anger face evolved to enhance cues of strength. *Evol Hum Behav*. doi:10.1016/j.evolhumbehav.2014.05.008

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. *Proc Roy B Sco*, *276*, 575–84. 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. *Evol Hum Behav, 33*, 121–129. doi:10.1016/j.evolhumbehav.2011.07.002

Skorska, M. N., Geniole, S. N., Vrysen, B. M., McCormick, C. M., & Bogaert, A. F. (in press). Facial structure predicts sexual orientation in both men and women. *Arch Sex Behav*.

Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychol Sci*, *21*, 349–54. doi:10.1177/0956797610362647

Számadó, S. (2008). How threat displays work: Species-specific fighting techniques, weaponry and proximity risk. *Anim Behav*, *76*, 1455–1463. doi:10.1016/j.anbehav.2008.07.010

Třebický, V., Fialová, J., Kleisner, K., Roberts, S. C., Little, A. C., & Havlíček, J. (2014). Further evidence for links between facial width-to-height ratio and fighting success: Commentary on Zilioli et al. (2014). *Aggress Behav*. doi: 10.1002/ab.21559

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

Undurraga, E. A. et al. (2010). Human’s cognitive ability to assess facial cues from photographs: a study of sexual selection in the Bolivian Amazon. *PLOS ONE, 5*, e11027. doi:10.1371/journal.pone.0011027

Van Staaden, M. J., Searcy, W. A. & Hanlon, R. T. (2011). *Signaling Aggression.* 1st edn. Elsevier Inc. doi:10.1016/B978-0-12-380858-5.00008-3

Vernon, R.J.W., Sutherland, C.A.M., Young, A.W., & Hartley, T.W. (2014). Modeling first impressions from highly variable facial images. *Proc Nat Acad Sci USA*, *111*, E3353-E3361. doi: 10.1073/pnas.1409860111

Welker, K. M., Goetz, S. M., Galicia, S., Liphardt, J., & Carré, J. M. (2014). An examination of the associations between facial structure, aggressive behavior, and performance in the 2010 World Cup Association Football Players. *Adapt Hum Behav Physiol*, 1-13. doi:10.1007/s40750-014-0003-3

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

Windhager, S., Schaefer, K., & Fink, B. (2011). Geometric morphometrics of male facial shape in relation to physical strength and perceived attractiveness, dominance, and masculinity. *Amer J Hum Biol, 23*, 805–14. doi:10.1002/ajhb.21219

Zebrowitz, L. A, Kikuchi, M. & Fellous, J. M. (2010). Facial resemblance to emotions: Group differences, impression effects, and race stereotypes. *J Pers Soc Psychol, 98*, 175–89. doi:10.1037/a0017990

Zilioli, S., Sell, A. N., Stirrat, M., Jagore, J., Vickerman, W., & Watson, N. V. (2014). Face of a fighter: Bizygomatic width as a cue of formidability. *Aggress Behav*, epub ahead of print, doi:10.1002/ab.21544

Rationale for Chapter 5

In Chapter 4, I provided additional evidence that the face ratio, and sensitivity to it, may be part of an evolved mechanism designed for the advertisement and assessment of threat potential. Participants' judgements of aggression were correlated with the face ratio even when men were bearded and, thus, it is possible that this metric was operational in shaping judgements of threat potential in our ancestral past. In Chapter 5, I again test predictions derived from the hypothesis that the face ratio, and sensitivity to it, is part of an evolved system designed for advertising and assessing threat potential. I synthesized all face ratio findings reported before December 31, 2014, by performing several meta-analyses. The mean weighted effects that I calculated were derived from studies involving participants belonging to many different nationalities (e.g., Canada, USA, Mexico, Germany, Czech Republic, China, Japan, countries within the United Kingdom). Based on the hypothesis that the face ratio and sensitivity to it may be part of an evolved threat-detection mechanism, it should be associated with behaviour and observer's judgements across these different countries. Therefore, I predicted that the face ratio would possess ecological validity (i.e., would share associations with threat-related behaviours) and that observers would utilize it when forming judgements about one's propensity to commit such behaviours (i.e., the ratio would share associations with threat-related judgements) when combining effect sizes from these studies that were conducted in various countries around the world.

Chapter 5: Evidence from meta-analyses of the facial width-to-height ratio as an evolved cue of threat

The published version of this chapter is:

Geniole, S.N., Denson, T., Dixson, B. Carré, J.M. & McCormick, C.M. (2015). Evidence from meta-analyses of the facial width-to-height ratio as an evolved cue of threat. *PLOS ONE*, *10*(7), e0132726. doi: 10.1371/journal.pone.0132726

Author Contribution: I was the primary investigator of this work, responsible for the majority of the study design, data analysis, and writing of the manuscript.

Introduction

Perceptual and sensory systems have evolved to detect threat (Blanchard, Griebel, Pobbe, & Blanchard, 2011). These systems are tuned to cues of formidability and aggressiveness in conspecifics, allowing for appropriate submissive or attack behaviours depending on the information conveyed by the cues (Arnott & Elwood, 2009; Maynard Smith & Harper, 2003; Parker, 1974). The rapid communication of rank, dominance, and fighting ability may curtail the escalation of agonistic encounters; there is much evidence that agonistic contests are settled more quickly and are less likely to be lethal when animals have visual exposure to their opponent before engaging in a contest than when they do not (e.g., in cichlids, *Cichlidae*; green swordtails, *Xiphophorus hellerii;* rainbow trout, *Oncorhynchus mykiss*; pigs, *Sus scrofa*; hamsters, *Mesocricetus brandti*, Arnott & Elwood, 2009). Although visual assessments likely depend on multiple cues of varying complexity (Arnott & Elwood, 2009), selection should favour conspicuous cues that are rapidly processed (Grueter, Isler, & Dixson, 2015).

In humans, the visual system is highly sensitive to, and quick to process, cues in the face such as identity, gender, age, and emotional expression (McGugin & Gauthier, 2013), which guide social interactions (Bruce & Young, 2012). Although emotional expressions account for much of this communication, static features in the face may also provide information such as formidability and aggressiveness; such static cues have been described in other species (e.g., black facial pattern of paper wasps, *Polistes dominulus*, Tibbetts & Lindsay, 2008). There is abundant evidence that humans form snap judgements of dominance and threat (e.g., aggressiveness, strength, fighting ability, Sell et al., 2009). Additionally, there is evidence that such judgements are accurate: people who were judged as more powerful reported being higher in assertiveness, social potency, aggressiveness, and power (Berry, 1991); those who were judged as stronger and better at fighting were physically stronger and reported fighting more frequently (Sell et al., 2009); and criminals who were judged as more violent were more likely to have been incarcerated for violent than for non-violent crimes than were criminals who were judged as less violent (Stillman, Maner, & Baumeister, 2010).

The facial width-to-height ratio (FWHR; the width of the face divided by the height of the upper face) may be an important static cue of threat; it is perceived rapidly (Carré, Morrissey, Mondloch, & McCormick, 2010), it is conspicuous even in bearded men (Geniole & McCormick, 2015), and it predicts men's aggressive behaviour both in and outside of the laboratory (Carré & McCormick, 2008; Goetz et al., 2013). Observers' estimates of aggression, dominance, and formidability are reliably correlated with the FWHR (e.g., Carré, McCormick, & Mondloch, 2009; Zilioli et al., 2015). Further, the FWHR is positively associated with dominance in non-human primates (Lefevre et al., 2014), and humans can accurately assess this trait in non-human primates (Kramer, King, & Ward, 2011; Kramer & Ward, 2012), suggesting that the FWHR, and sensitivity to it, may be part of an evolved cuing system in human and non-human primates. Nevertheless, the reliability of these relationships (e.g., Gómez-Valdés et al., 2013; Özener, 2012) and the report of a larger FWHR in men than in women (e.g., Kramer, Jones, & Ward, 2012; Lefevre et al., 2012) have been questioned (see Figure 5-1 for examples of faces with relatively low and with relatively high FWHRs).

The current meta-analytic review

The abundance of research on the FWHR since its first report in humans (Weston, Friday, & Liò, 2007) permits an assessment of the reliability and magnitude of these relationships. Although a meta-analysis on the FWHR was recently published (Haselhuhn, Ormiston, & Wong, 2015), the scope of that analysis was limited to characterizing the relationship between the FWHR and aggression among men only. Here, we systematically review a greater body of FWHR research and we use meta-analyses to investigate whether this metric: (1) is sexually dimorphic and cues judgements related to masculinity and femininity; (2) cues judgements of threat and dominance across several domains; (3) is an accurate index of these characteristics and behaviours in both men and women, (4) is associated with attractiveness judgements, and; (5) body mass index. In so doing, we provide a more definitive test of the hypothesis that the FWHR is part of an evolved cueing system of intra-sexual threat, dominance, and aggressiveness in men, akin to those in other species (e.g., Tibbetts & Lindsay, 2008).

Methods

We identified all peer-reviewed and published or in-press manuscripts written in English that contained effect sizes related to the FWHR by using the search term "facial width-to-height ratio" in Google Scholar and by searching for citations of Weston, Friday, and Liò (2007), the first article published on the FWHR (our search ended December 31st, 2014). We also included effect sizes from four separate manuscripts that were submitted for review by authors of the current manuscript (Denson, unpublished manuscript; Yang, Chao, Fabiansson, & Denson, unpublished manuscript); two of which have been accepted since (Geniole & McCormick, 2015; Welker, Goetz, & Carré, 2015) and one manuscript of the authors that included the term "facial width-to-height ratio"

(Geniole & McCormick, 2013) but was not detected by Google Scholar. This strategy identified 63 peer-reviewed manuscripts. Effect sizes from seven of these manuscripts were not used in any of the meta-analyses, however, because the authors either did not conduct analyses that were relevant to our research questions (Abel, Kruger, & Dai, 2013; Bryan, Perona, & Adolphs, 2012; Kleiman & Rule, 2013), involved non-human primates (Carré, 2014; Lefevre et al., 2014; V. Wilson et al., 2014), or used faces intentionally posed in non-neutral expressions (Marsh, Cardinale, Chentsova-Dutton, Grossman, & Krumpos, 2014). Therefore, analyses were conducted on effect sizes extracted from a total of 56 manuscripts.

We used an effect size determination program (D. B. Wilson, 2001) and formulas provided in Bonett (Bonett, 2007) and Tabachnick and Fidell (Tabachnick & Fidell, 2007) to convert effect sizes to either a Pearson product moment correlation (*r*) or to a standardized mean difference (*d*). When estimating effect sizes from studies using multilevel modelling or binary logistic regression, we converted the χ^2 values from the individual predictors to *r* or *d* values using Wilson's (D. B. Wilson, 2001) effect size determination program, or we computed a *t* value by dividing the coefficient by the standard error of the coefficient (as in Schulz, Cowan, & Cowan, 2006), and converted this value to an *r* or *d* value using Wilson's program (D. B. Wilson, 2001). When standardized coefficients (*ß* weights) were provided instead of *r* values, we used *ß* weights as direct estimates of *r* values given their equivalence when a variable is entered as the sole predictor in a regression, and their strong correlation when the variable of interest is entered along with several other simultaneous predictors in a regression (Peterson & Brown, 2005). When η^2 values were provided, we used the square root of these values as an estimate of the *r* effect size. Three of the authors coded all effect sizes; discrepancies were resolved through discussion. Additional detail regarding data extraction and effect size conversions are in the Supplementary Materials and Methods section.

For meta-analyses involving the *d* effect size values, the *d*s were adjusted to correct for small sample size bias [$d(1-(3/(4N-9)))$] (Hedges & Olkin, 1985) and were weighted by the inverse variance ($1/se^2$) before calculation of the mean weighted effect size. Therefore, all *d* (for individual effect sizes) and $\bar{d}$ (for mean weighted effect sizes) values are presented in the adjusted, unbiased form in tables, figures, and text unless otherwise stated. As recommended (Lipsey & Wilson, 2001), for meta-analyses involving the *r* effect size values, the *r*s were transformed to Fisher z correlations and weighted by the inverse variance ($N-3$) before calculating the mean weighted effect size. For ease of interpretation, however, these Fisher z estimates were then transformed back into their standard *r* (for individual effect sizes) or their $\bar{r}$ (for mean weighted effect sizes) form when presented in tables, figures, and text.

The data were analyzed using SPSS macros with random-effects models (D. B. Wilson, 2001). The macro “MEANES” was used to determine the mean weighted effect sizes; the macro “METAF” was used to test individual moderators with two levels. When an individual moderator with two levels was significant, the file was split by the moderating variable and the macro “MEANES” was used to determine the mean weighted effect size within each level or subgroup. The macro “METAREG” was used to test moderators with continuous values, or was used to test multiple moderators (with discrete levels, or with continuous values) simultaneously (e.g., to test the effect of one moderator, statistically controlling for the other moderators). Although we present results separately for each subgroup when a moderator with two levels was tested independently and found to be significant, we only provide *B* weights when the moderator had continuous rather than ordinal values or when it was tested simultaneously with other moderators. The *B* weights can be used to determine the extent to which the mean weighted effect size changes with each unit change of the moderator variable (controlling statistically for any other moderators that may be included in the model). Therefore, if a moderator had a *B* weight of .20, the strength of the mean weighted effect size increases by .20 with a one unit increase in the moderator. Similarly, if the moderator involved two levels, it would suggest that the relationship within one level differed .20 from the relationship within the other level. All moderators were tested separately (without other moderators in the model) unless otherwise specified.

Our meta-analysis on the relationship between the FWHR and threat differs from that of Haselhuhn and colleagues (Haselhuhn et al., 2015) in that we included a broader array of behaviours (e.g., prejudice, financial misreporting) related to threat, and also investigated the association in both men and women. Although there are discrepancies in the definition of "threat" in the literature, we use the term according to its definition in the Merriam-Webster dictionary (http://www.merriam-webster.com/dictionary/threat): "someone or something that could cause trouble, harm, etc." Because this definition is broad, it captures many related yet distinct behaviours (e.g., aggression, prejudice, deception). We therefore conducted moderator analyses to examine whether the association between the FWHR and threat differs in strength depending on the type of threat; we distinguished between the most commonly investigated type of threat, aggressive behaviour, and other selfish and pejorative behaviours.

Our analysis also differs from that of Haselhuhn and colleagues (2015) in that we estimated the means and standard deviations from Figure 2-B of Gómez-Valdés and colleagues (2013) rather than assume the relationship between the FWHR and threat behaviour was r = .00. Compared to the analysis of Haselhuhn and colleagues (2015), which included 4141 men from 18 samples, our analysis of men included 4573 participants (and 30 male dyads) from 23 samples. Again, the samples included in our analysis were derived from studies involving a broader array of behaviours related to threat than those included in Haselhuhn and colleagues' (2015) analysis. We excluded a study that investigated death by contact violence (Stirrat, Stulp, & Pollet, 2012) because this study examined aggression towards, rather than aggression perpetrated by, the individual; this study, however, was in the meta-analysis of Haselhuhn and colleagues (2015). We also used a random-effects model to analyse all data rather than use a fixed-effects model, which was used by Haselhuhn and colleagues (2015).

When deciding which effect sizes to extract for examining the relationship between the FWHR and dominance, we referred to the definition of dominance in the Merriam-Webster dictionary, one's relative position within a social hierarchy (http://www.merriam-webster.com/dictionary/dominance), as well as questionnaire measures of dominance and prestige [e.g., "I do NOT have a forceful or dominant personality" (reversed); "I try to control others rather than permit them to control me"; "I often try to get my own way regardless of what others may want"; "Others always expect me to be successful"; (Cheng, Tracy, & Henrich, 2010); "Have a strong need for power";

(Goldberg et al., 2006)]. The analysis included effect sizes related to self-perceived, other-perceived, or objectively determined prestige, forcefulness, inflexibility, competitiveness, military rank, sense of power, and achievement drive.

For moderator analyses, we extracted information related to the nationality and mean age of the samples and, if a study involved observers' perceptions, the number of observers and the mean age and nationality of the faces used as stimuli. We also extracted information on the measurement of the FWHR (2D photos, 3D scans, etc; see Supplementary Materials and Methods for additional notes regarding moderators) for the analysis of sex differences in the size of the FWHR. Although we provide funnel plots for each analysis involving 10 or more effect sizes (Sterne & Egger, 2001) (see supplementary Figure 5-1), we caution that asymmetry in the plots may arise for a number of reasons other than, or in addition to, publication bias (e.g., true heterogeneity in the effect size, poor methodological design of smaller studies, chance; reviewed in Egger, Davey Smith, Schneider, & Minder, 1997). When there were significant moderators of effect sizes and the moderators were discrete variables (rather than continuous), we provide the funnel plots within each subgroup (unless the subgroup involved a small number of studies, $k < 10$). We also provide a fail-safe n (Lipsey & Wilson, 2001; Orwin, 1983) for each significant effect ($p \leq .05$), indicating the number of additional studies with null effects that would have to be added to the analysis to make the magnitude of the mean weighted effect size trivial, $\bar{r} = .01$ or $\bar{d} = 0.01$. We visually inspected the funnel plots for any potential outliers. For each meta-analysis, we report the number of samples that were included (k).

Results

Table 5-1 provides a summary of the results.

Are men's FWHRs larger than women's?

Studies were included in the analysis if they reported statistics comparing the FWHR of men and women, or descriptive statistics regarding the size of the FWHR (means, *SD*) for the sexes separately. With these inclusion criteria, 19 of the 56 manuscripts were included in the analysis. Effect sizes were extracted from 32 samples involving 6113 men and 4740 women (Table S5-1) (M_{age} = 25.19 years; range: 18.98 - 83). Men had slightly larger FWHRs than did women ($k = 32$, $\bar{d} = 0.11$, *95% CI* = 0.03 to 0.20, $p = .009$; $Q_{31} = 110.49$, $p < .0001$; fail-safe $n = 320$), even after removing the largest outlying effect size ($k = 31$, $\bar{d} = 0.08$, *95% CI* = 0.003 to 0.16, $p = .04$; $Q_{30} = 80.32$, $p < .0001$; fail-safe $n = 310$). Neither age, measurement type (2D photographs vs other), nor nationality (North American vs other) moderated the effect (*p*s > .21).

Are larger FWHRs perceived as more masculine than are smaller FWHRs?

Studies were included in the analysis if they reported statistics examining the association between the FWHR and judgements of masculinity or of femininity (femininity correlations were reversed for the analysis). With these inclusion criteria, six of the 56 manuscripts were included in the analysis using correlational design and/or a continuum of faces with un-manipulated FWHRs.

Studies using a correlational design and/or a continuum of faces with un-manipulated FWHRs. Effect sizes were extracted from 12 samples (Table S5-2), which involved a total of 139 male observers from 9 of the samples and 200 female observers from 11 of the samples (M_{age} = 25.70). The stimuli included 425 male faces from 10 of the samples and 62 female faces from two of the samples (M_{age} = 19.42). The mean weighted relationship between the FWHR and perceptions of masculinity was positive and significant (k = 12, $\bar{r}$ = .30, *95% CI* = .18 to .42, p < .0001; Q_{11} = 19.53, p = .05; fail-safe n = 348). Sex of the stimuli moderated the effect (k = 12, Q_1 = 4.62, p = .03), with stronger effects in male (k = 10, $\bar{r}$ = .35, *95% CI* = .23 to .47, p < .0001; Q_9 = 14.63, p = .10; fail-safe n = 340) than in female faces (k = 2, $\bar{r}$ = -.01, *95% CI* = -.26 to .25, p = .97; Q_1 = 0.11, p = .74). Although the relationship between the FWHR and perceptions of masculinity/femininity differed for male and female faces, there were only two samples from which the estimate for female faces was derived. Further, there were only 31 unique female facial identities used in the analysis. Therefore, future studies would benefit from examining this potential moderating factor using a larger set of unique female faces as stimuli.

Within the samples using male faces as stimuli, neither the number of observers, age of observers, nor age of stimuli moderated the effect (*p*s > .21). The percentage of male observers was a significant moderator (k = 10, B = -.004, p = .003), but the effect was driven by one effect size (Sanchez-pages, Rodriguez-ruiz, & Turiegano, 2014). After its removal, the moderator was not significant (k = 9, B = .000, p = .91). Among studies using male faces as stimuli, the most frequently used stimuli set (Carré et al., 2009) did not produce stronger effect sizes than did studies involving other stimuli sets (p = .59).

Studies using faces with manipulated FWHRs. No studies to date examined perceptions of masculinity between two versions of a face manipulated to have smaller versus larger FWHRs.

Does the FWHR predict threatening and dominant behaviour?

Threat. Nineteen of the 56 manuscripts met the inclusion criteria for the analysis of the association between the FWHR and threat behaviour (selfish, pejorative, and aggressive behaviour). Effect sizes were extracted from 32 samples (Table S5-3). There was a total of 4573 men (and 30 male dyads) from 23 of the samples and 634 women from 9 of the samples (M_{age} = 21.77 years; range: 18.98 – 28). The FWHR predicted threat behaviour ($k = 32$, $\bar{r} = .12$, *95% CI* = .07 to .17, $p < .0001$; fail-safe $n = 352$) despite the presence of one apparent outlier (Gómez-Valdés et al., 2013), which was the only effect size with confidence intervals that did not overlap with those of the mean weighted effect size ($r = -.34$, *95% CI* = -.47 to -.20) (Figure 5-2). Note that the outlying effect size was computed by comparing the FWHRs of the general population to the weighted mean of three criminal groups (prosecuted for committing homicide, robbery, or other minor faults, see Figure 2-B of Gómez-Valdés et al., 2013, and footnote in Table S5-3). Nevertheless, the FWHR may be related to socioeconomic success (given its links with performance in economic negotiations and business; see analysis below), which is known to predict criminality (e.g., Levine, 2011), thus representing a potential suppression effect. In support of this possible suppression effect, when we minimize the influence of socioeconomic status by making comparisons within the criminal group [comparing the group prosecuted for committing homicide ($n = 58$, mean = 1.838, SD = 0.118) to the weighted mean of those prosecuted for committing robbery and other minor faults (total $n = 49$, weighted mean = 1.803, pooled SD = 0.111)] the effect size becomes positive ($r = .15$) and more consistent with the mean weighted effect size reported for men.

Excluding this outlying effect size increased the mean weighted effect size, tightened the confidence interval (k = 31, $\bar{r}$ = .13, *95% CI* = .09 to .17, p < .0001; fail-safe n = 372), and reduced the heterogeneity (outlier included: Q_{31} = 83.60, p < .0001; outlier excluded: Q_{30} = 50.96, p = .01). Because this analysis also involved effect sizes from studies that investigated fighting ability (Třebický et al., 2015; Zilioli et al., 2015), which involves a combination of aggressiveness and athletic ability, we re-ran the analysis without these studies included. The mean weighted effect size was unchanged although the confidence intervals became slightly wider (k = 29, $\bar{r}$ = .13, *95% CI* = .09 to .18, p < .0001; fail-safe n = 360). This analysis also involved some effect sizes that may have come from overlapping samples in different manuscripts (UFC fight performance: Třebický et al., 2015; Zilioli et al., 2015; penalty minutes of players from the National Hockey League: Carré & McCormick, 2008; Deaner, Goetz, Shattuck, & Schnotala, 2012; Goetz et al., 2013). When we included only the effects size from the largest sample of the overlapping studies involving UFC fighters (Zilioli et al., 2015) and of the overlapping studies involving hockey players (Goetz et al., 2013), the mean weighted effect size was unchanged (k = 28, $\bar{r}$ = .13, *95% CI* = .08 to .18, p < .0001; fail-safe n = 348).

Sex interacted with the FWHR (k = 31, Q_1= 5.37, p = .02); the relationship was significant only in men (k = 22, $\bar{r}$ = .16, *95% CI* = .11 to .21, p < .0001; Q_{21} = 43.15, p = .003; fail-safe n = 330) (women: k = 9, $\bar{r}$ = .04, *95% CI* = -.04 to .12, p = .34; Q_8 = 4.91, p = .77; Figure 5-2). Note that the interaction involving sex was marginally significant when the outlying effect size from Gómez-Valdés and colleagues (2013) was included in the analysis (k = 32, Q_1 = 3. 53, p = .06). The type of threat (aggressive vs selfish and pejorative) was not a moderator (k = 22, Q_1 = 2.38, p = .12), with the FWHR predicting both aggressive (k = 12, $\bar{r}$ = .13, *95% CI* = .07 to .18, p < .0001; Q_{11} = 19.72, p = .05; fail-safe n = 144) and selfish/pejorative behaviour (k = 10, $\bar{r}$ = .25, *95% CI* = .14 to .35, p < .0001; Q_9 = 21.78, p = .01; fail-safe n = 240) in men. The type of measure (self-report vs behavioural) did not moderate the relationship (k = 22, Q_1 = 0.18, p = .67) (behavioural measures: k = 17, $\bar{r}$ = .17, *95% CI* = .11 to .22, p < .0001; Q_{16} = 36.96, p = .002; fail-safe n = 272; self-report measures: k = 5, $\bar{r}$ = .14, *95% CI* = -.01 to .29, p = .07; Q_4 = 6.13, p = .19). Nationality (North American vs other) moderated the relationship (k = 22, Q_1 = 6.65, p = .01); the FWHR shared stronger relationships with threat behaviour when effect sizes were derived from North American (k = 9, $\bar{r}$ = .25, *95% CI* = .17 to .33, p < .0001; Q_8 = 7.85, p = .45; fail-safe n = 216) than from other samples (k = 13, $\bar{r}$ = .12, *95% CI* = .07 to .17, p < .0001; Q_{12} = 23.05, p = .03; fail-safe n = 143), although the mean weighted effect was significant irrespective of the nationality. Age did not moderate the relationship between the FWHR and threat behaviour (p = .42).

Dominance. Ten of the 56 manuscripts met the inclusion criteria for the analysis. Effect sizes were derived from 17 samples (Table S5-4), with a total of 1426 men (and 30 male dyads) from 11 of the samples and 287 women from 6 of the samples (M_{age} = 22.04 years; range: 18.98 – 33.61). All studies involved subjective measures of dominance (either self-report or, for a study involving previous presidents, Lewis, Lefevre, & Bates, 2012, inferred dominance) except one (Loehr & Hara, 2013) that involved the relationship between the FWHR and military rank of Finnish soldiers at the start of World War II. The relationship between the FWHR and dominance was positive and significant (k = 17, $\bar{r}$ = .10, *95% CI* = .002 to .20, p = .05; Q_{16} = 45.16, p = .0001; fail-safe n = 153). When only studies that involved self-reported or inferred measures of dominance were included, the confidence interval was tighter and the distribution of effect sizes was no longer heterogeneous (k = 16, $\bar{r}$ = .12, *95% CI* = .05 to .18, p = .0005; Q_{15} = 14.68, p = .47; fail-safe n = 176), likely because the study of Finnish soldiers (Loehr & Hara, 2013) produced the only effect size with confidence intervals that did not overlap with those of the mean weighted effect size. Sex did not moderate the relationship (k = 16, Q_1 = 0.20, p = .65) (men: k = 10, $\bar{r}$ = .14, *95% CI* = .04 to .24, p = .008; Q_9 = 13.59, p = .14; fail-safe n = 130) (women: k = 6, $\bar{r}$ = .09, *95% CI* = -.03 to .21, p = .12; Q_5 = 0.91, p = .97). Neither nationality (North American vs other) nor age moderated the relationship between the FWHR and dominance (ps > .29).

We included the studies of business-related outcomes (any effect sizes related to negotiation abilities, business position) and sports performance (any effect sizes related to wins and indices of successful performance in sports, e.g., assists, goals) as additional indices of dominance. With the inclusion criteria, 4 of the 56 manuscripts were included in the analysis on business-related outcomes. The analysis included effect sizes from 6 samples (Table S5-5) involving a total of 241 men (and 87 male dyads and 86 male groups) (M_{age} = 27.14). The FWHR predicted success in business, marginally (k = 6, $\bar{r}$ = .22, *95% CI* = -.04 to .46, p = .09; Q_5 = 29.23, p < .0001). The association was negative in only one study (Study 3 of Haselhuhn, Wong, Ormiston, Inesi, & Galinsky, 2014), which was similar to other studies included in the analysis in that it examined the ability to negotiate, but differed from other studies in that it assessed the ability to negotiate legitimately (within the rules of the bargaining exercise). This effect size may have been opposite to the other effect sizes because it represents a measure of bargaining within the rules of the bargaining game. Nevertheless, as the analyses above suggest, men with larger FWHRs are more antisocial than those with smaller FWHRs and this effect may thus be driven by an increased likelihood of "cheating" in the task to achieve the goal. When this effect size was excluded, the mean weighted effect became significant and had narrower confidence intervals (k = 5, $\bar{r}$ = .32, *95% CI* = .12 to .50, p = .002; Q_4 = 13.61, p = .009; fail-safe n = 155). Neither nationality (North American vs other) nor age moderated the relationship (ps > .60).

Four of the 56 manuscripts met the inclusion criteria for the analysis of sports performance. The analysis included effect sizes from 4 samples (Table S5-6) involving a total of 1401 men (M_{age} = 29.34). The FWHR predicted sports performance (k = 4, $\bar{r}$ = .10, *95% CI* = .005 to .19, p = .04; Q_3 = 6.57, p = .09; fail-safe n = 36). One of the samples (Welker, Goetz, Galicia, Liphardt, & Carré, 2014) included a measure of performance in soccer players (the average of the associations between the FWHR and assists and between the FWHR and goals). The authors performed analyses controlling for player position (defender, midfielder, forward) and also within each player position. Because forwards have more opportunities to score goals and make more assists than do midfielders and defenders (Welker et al., 2014), we also examined the mean weighted association between the FWHR and sports performance when this subsample of forwards (n = 211) was used instead of the entire sample. The mean weighted effect size from this analysis was stronger, and the heterogeneity was reduced (k = 4, $\bar{r}$ = .15, *95% CI* = .08 to .22, p = .0001; Q_3 = 0.35, p = .95; fail-safe n = 56).

Are perceptions of threat and dominance associated with the FWHR?

Studies of threat using a correlational design and/or a continuum of faces with un-manipulated FWHRs. For the analysis of perceived threat, we included any studies that reported statistical analyses on the relationship between the FWHR and threat-related judgements (see definition of threat in methods). These judgements included those of aggressiveness, untrustworthiness, formidability (strength, toughness, fighting ability, physical power), and prejudice. With these inclusion criteria, 18 of the 56 manuscripts were included in the analysis of studies that used a correlational design and/or a continuum of faces with un-manipulated FWHRs. Effect sizes were extracted from 38 samples (Table S5-7) involving a total of 779 male observers from 36 of the samples and 1313 female observers from all 38 of the samples (M_{age} = 21.57) (see Figure 5-2). The stimuli included 1679 male faces from 36 of the samples and 72 female faces from three of the samples (M_{age} = 22.64). The FWHR predicted perceptions of threat (k = 38, $\bar{r}$ = .48, *95% CI* = .41 to .55, p < .0001; Q_{37} = 125.43, p < .0001; fail-safe n = 1786). When we removed the largest effect size (r = .81, Study 3 of Hehman, Leitner, & Freeman, 2014), the strength of the association decreased slightly (k = 37, $\bar{r}$ = .46, *95% CI* = .39 to .53, p < .0001; Q_{36} = 96.71, p < .0001; fail-safe n = 1739).

Because a cluster of the variables we investigated as moderators were correlated with one another (all *r*s > .28, *p*s < .10), we entered them as simultaneous moderators [number of observers, nationality of the observers (North American vs other), nationality of the faces used as stimuli (North American vs other), age of the faces used as stimuli (younger than 25 vs older than 25)]. Only the age of the stimuli emerged as a significant moderator (k = 25, B = -.28, p = .01) (all other *p*s > .34), with perceptions of threat sharing stronger links with the FWHR of younger compared with older individuals. The percentage of male observers, the age of observers, whether the stimuli included female faces, and the type of the judgement (judgements of only aggression vs other), were not significant moderators (all $p > .10$). When these variables were entered as simultaneous moderators along with the age of the stimuli faces, the only significant moderators were age of the stimuli faces ($k = 26$, $B = -.40$, $p < .0001$) and judgement type ($k = 26$, $B = .15$, $p = .01$); the FWHR predicted perceptions of threat more strongly in younger than in older faces, and when participants' judgements were of aggression on its own compared with when other threat judgements were involved.

We also examined whether there were any differences in the strength of the association between the FWHR and judgements of threat when effect sizes were obtained from studies using the most common stimuli set (24 male faces from Carré et al., 2009) compared to other stimuli sets. Studies using the stimuli set from Carré and colleagues (2009) ($k = 11$, $\bar{r} = .61$, *95% CI* = .52 to .68, $p < .0001$; $Q_{10} = 7.57$, $p = .67$; fail-safe n = 660) produced stronger effect sizes ($k = 37$, $Q_1 = 9.02$, $p = .003$) than did studies using other stimuli sets ($k = 26$, $\bar{r} = .40$, *95% CI* = .32 to .48, $p < .0001$; $Q_{25} = 64.82$, $p < .0001$; fail-safe n = 1014). Because the Carré and colleagues (2009) stimuli set involved younger faces, was more often used to assess perceptions of aggression, and was more often rated by North American observers, compared with other stimuli sets (all *r*s > .30), we examined whether these three variables explained its stronger associations. When these three moderators (age of faces used as stimuli, nationality of the observers, judgement type) were entered simultaneously with the type of stimuli set (Carré and colleagues vs other), only the age of the faces emerged as a significant moderator ($k = 25$, $B = -.32$, p = .0007; all other *p*s >.05), suggesting that these variables may, in part, explain the stronger associations obtained when studies used the Carré and colleagues' stimuli set compared to other stimuli sets.

Studies of threat using faces with manipulated FWHRs. Six of the 56 manuscripts met the inclusion criteria for the analysis of studies involving faces with manipulated FWHRs. Effect sizes were extracted from 11 samples (Table S5-8) involving a total of 467 male observers from all 11 samples and 9135 female observers from all 11 of the samples as well (M_{age} = 25.78). Faces with larger FWHRs were rated as more threatening than those with smaller FWHRs, but the difference missed statistical significance (k = 11, $\bar{d}$ = 0.42, *95% CI* = -0.02 to 0.86, p = .06; Q_{10} = 155.65, p < .0001). The heterogeneity was driven by one outlying effect size (Study 1 of Hehman et al., 2014), which was removed from subsequent analyses, and three effect sizes that were in a direction opposite to that of the other seven effect sizes. The three negative effect sizes were derived from studies using small stimulus sets (one manipulated stimulus face, Bashir & Rule, 2014; two male and two female manipulated stimulus faces, Wang, Geigel, & Herbert, 2013), which may have obscured the relationship between the FWHR and perceptions of threat. Consistent with this possibility, the number of base stimulus images or composites moderated the strength of the effect size (k = 10, B = .09, p < .0001), such that studies that utilized more base stimulus images or composites produced larger effect sizes than those that used fewer. In addition to the size of the stimulus set, it is also possible that two of the three effect sizes were in a direction opposite to that of the rest of the studies because they came from a study that involved the use of avatars that were caricatured rather than realistic in appearance (see Figure 1 of Wang et al., 2013). Further, for the other negative effect size, the manipulation of the FWHR may have incidentally exaggerated the lower jaw and increased perceptions of adiposity, which may have influenced the judgements (see Figure 3 of Bashir & Rule, 2014).

After excluding the studies that used only one or two base images, faces manipulated to have larger FWHRs were perceived as significantly more threatening than those manipulated to have smaller FWHRs, and heterogeneity was reduced (k = 7, $\bar{d}$ = 0.41, *95% CI* = 0.29 to 0.53, p < .0001; Q_6 = 5.12, p = .53; fail-safe n = 287). In this smaller sample of effect sizes (k = 7), neither threat type (aggression vs other), sex of stimuli, percentage of male observers, nationality of the stimuli (North America vs other), nor nationality of the observers (UK vs other) (ps > .12) moderated the effect.

Studies of dominance using a correlational design and/or a continuum of faces with un-manipulated FWHRs. For the analysis of perceived dominance, we included any studies that reported statistical analyses on the relationship between the FWHR and dominance-related judgements (see definition of dominance in methods). Seven of the 56 manuscripts met the inclusion criteria for the analysis involving studies that used a correlational design and/or a continuum of faces with un-manipulated FWHRs. Effect sizes were extracted from 8 samples (Table S5-9) involving a total of 107 male observers from all eight of the samples and 153 female observers from all eight of the samples (M_{age} = 24.88). The stimuli included 461 male faces from seven of the samples and 202 female faces from three of the samples (M_{age} = 28.89). The FWHR predicted perceptions of dominance (k = 8, $\bar{r}$ = .29, *95% CI* = .10 to .47, p = .004; Q_7 = 45.08, p < .0001; fail-safe n = 224). Nevertheless, one effect size seemed to be driving the effect (Study 3 of Hehman et al., 2014); after excluding the study, the effect size decreased in magnitude but was still positive and significant (k = 7, $\bar{r}$ = .20, *95% CI* = .06

to .34, p = .007; Q_6 = 18.70, p = .005; fail-safe n = 196). The number of observers, percentage of male observers, age of observers, nationality of observers (North America vs other), age of stimulus faces, and nationality of stimulus faces (North America vs other) did not moderate the relationship between the FWHR and perceived dominance (ps > .19). The relationship between the FWHR and perception of dominance, however, was marginally stronger (k = 7, Q_1 = 3.62, p = .06) when stimuli sets were exclusively male faces (k = 4, $\bar{r}$ = .30, *95% CI* = .19 to .40, p < .0001; Q_3 = 2.53, p = .47, fail-safe n = 140) than when they were not (k = 3, $\bar{r}$ = .06, *95% CI* = -.19 to .31, p = .64; Q_2 = 9.07, p = .01).

Studies of dominance using faces with manipulated FWHRs. Only one study that manipulated the FWHR and investigated changes in perceptions of dominance fit the inclusion criteria (Bashir & Rule, 2014). The study manipulated a single male face to have a larger versus a smaller FWHR and reported significantly higher ratings of dominance for the version of the face with the larger than the smaller FWHR (*unadjusted* d = 0.61, 52 observers). This effect was not included in any of the meta-analyses.

Is the FWHR associated with perceived attractiveness?

Studies using a correlational design and/or a continuum of faces with un-manipulated FWHRs. For the analysis on perceived attractiveness, we included any studies that reported statistical analyses on the relationship between the FWHR and attractiveness-related judgements (attractiveness, short-term or long-term desirability as a romantic partner). Nine of the 56 manuscripts met the inclusion criteria for the analysis involving studies that used a correlational design and/or a continuum of faces with un-manipulated FWHRs. Effect sizes were extracted from 14 samples (Table S5-10) involving a total of 106 male observers from 10 of the samples and 229 female observers from all 14 of the samples (M_{age} = 24.81). The stimuli included 659 male faces from 12 of the samples and 62 female faces from two of the samples (M_{age} = 22.84). The relationship between the FWHR and perceptions of attractiveness was negative and significant (k = 14, $\bar{r}$ = -.26, *95% CI* = -.40 to -.10, p = .001; Q_{13} = 50.69, p < .0001, fail-safe n = 350). Neither the number of observers, age of observers, age of the stimuli, sex of the stimuli, the nationality of the stimuli (North American vs other), nor the nationality of the observers (North American vs other) moderated the effect (all *p*s > .15). The percentage of male observers, however, did moderate the effect (k = 14, B = .008, p = .01); the negative relationship between the FWHR and judgements of attractiveness was stronger when the sample had a greater proportion of women than men, suggesting that faces with larger FWHRs may be especially unattractive to female observers. The strength of the mean weighted effect size did not differ between studies using the most frequently used stimuli set, that of Carré and colleagues (2009), and other stimuli sets (p = .65).

Studies using faces with manipulated FWHRs. Only one study that manipulated the FWHR and investigated changes in perceptions of attractiveness met the inclusion criteria (Bashir & Rule, 2014). The study manipulated a single male face to have a larger versus a smaller FWHR and reported no significant differences between the ratings of attractiveness for the two versions of the faces (*unadjusted d* = 0.06, 55 observers). This effect was not included in any of the meta-analyses.

Is the FWHR associated with Body Mass Index (BMI)?

For the analysis on the FWHR and BMI, we only included studies that reported statistical analyses on the relationship between the FWHR and BMI; we did not include associations with other indices of size or adiposity. BMI may mediate the relationship between the FWHR and behaviour. Nine of the 56 manuscripts met the inclusion criteria for the analysis. Effect sizes were extracted from 22 samples (Table S5-11): There was a total of 1479 men from 16 of the samples and 1009 women from 11 of the samples (some samples involved both male and female participants and did not report results separately for men and women) (M_{age} = 25.19 years; range: 19.6 - 83). The mean weighted relationship between the FWHR and BMI was positive and significant ($k = 22$, $\bar{r} = .31$, *95% CI* = .26 to .36, $p < .0001$; $Q_{21} = 34.62$, $p = .03$, fail-safe $n = 660$). Neither nationality (UK vs other), sex, nor age moderated the relationship between the FWHR and BMI (*p*s > .28).

Examination of funnel plots

See supplementary Figure 5-1 for funnel plots. The funnel plots indicate that the distribution of effect sizes for most of the meta-analyses were symmetrical, which suggests that the estimates of the mean weighted effect sizes were not likely to be biased. One distribution of effect sizes that does appear asymmetrical, however, is that of the relationship between the FWHR and threat, with many smaller samples producing larger positive effect sizes. Although the fail-safe *n*s associated with these analyses indicate that the relationship was robust, the skew in the effect sizes suggests that the estimate for the mean weighted effect size for the FWHR-threat relationship may have been biased by the results of these smaller studies.

Discussion

Our meta-analyses addressed many outstanding discrepancies in the literature on the FWHR, and confirm its relationship with threat and dominant behaviour (a robust, albeit small, effect size) and with observers' judgements of these traits (robust, and larger effect sizes)[19]. Studies of the FWHR were propelled by Weston and colleagues' (2007) report that this metric was sexually dimorphic (men > women), independent of body size, and emerged at puberty coincident with the rise in androgens. Despite several failures to replicate the sex difference (Kramer et al., 2012; Lefevre et al., 2012), our meta-analysis revealed a small but significant sex difference in the FWHR, with men's FWHRs slightly larger than women's. Further, the meta-analysis indicated a positive association between judgements of masculinity and the FWHR in men. Although the independence of the FWHR to allometric scaling has not been tested further, studies have investigated the relationship between body mass index (BMI) and the FWHR (e.g., Coetzee, Chen, Perrett, & Stephen, 2010; Mayew, 2013), which our meta-analysis indicated was moderately associated with the FWHR in both sexes. Although this relationship with BMI may explain some of the association between the FWHR and behaviour (Mayew,

[19]Is there also evidence for a decreasing effect size in face ratio research? I tested this possibility by examining publication year (see publication years in the Supplementary Tables) as a moderator of the relationship between face ratio and threat. Publication year was a significant moderator, with the strength of the association between the face ratio and threat behaviour decreasing across subsequent years (k = 22, B = -0.03, 95% CI = -0.06 to -0.007, p = .01). This effect is likely driven by initial, small sample studies which required larger effect sizes to reach statistical significance. Publication year shared some association with sample size, albeit non-significant (r = .30, p = .17). When entered as simultaneous moderators, however, only the sample size emerged as a significant predictor (k = 22, B = -0.0002, 95% CI = -0.0003 to -0.0001, p = .0001; publication year: p = .10). Therefore, the size of the relationship between the face ratio and threat behaviour decreases over time, likely because initial studies involved smaller samples for which larger effect sizes were required to meet statistical significance. With larger samples, the effect size appears to approach a lower value more consistent with the mean weighted effect size. Publication year was also a significant moderator of the relationship between the face ratio and perceptions of threat, with the strength of the association decreasing across subsequent years (k = 37, B = -0.07, 95% CI = -0.11 to -0.02, p = .008). Sample sizes increased across subsequent publication years (r = .48, p = .003) and explained the association between publication year and the effect size: when both sample size and publication year were entered as simultaneous moderators, only sample size was significant (k = 37, B = -0.004, 95% CI = -0.005 to -0.002, p < .0001; publication year: p = .70). Therefore, the relationships between the face ratio and threat behaviour and between the face ratio and judgements of threat appear to be getting smaller across the years as a consequence of researchers using larger samples that better approximate the true effect sizes, which appear to be smaller than suggested by the initial studies on this facial metric.

2013), in several studies the relationships were similar when controlling for BMI and when BMI was not controlled (e.g., Lefevre, Lewis, Perrett, & Penke, 2013; Welker et al., 2015, 2014; Zilioli et al., 2015).

Weston and colleagues (2007) speculated that sexual dimorphism in the FWHR evolved via female choice as an attractive trait. Although studies have reported positive associations between the FWHR and male reproductive success (e.g., Gómez-Valdés et al., 2013; Loehr & Hara, 2013), the meta-analysis found that wider-faced men are judged as less attractive, especially by women, than are narrow-faced men. However, body size and androgen-dependent traits also function in intra-sexual competition (Puts, 2010). Indeed, our meta-analyses found that men with relatively larger FWHRs behaved in more threatening ways and described themselves as more aggressive, uncooperative, and prejudiced than did men with smaller FWHRs. Further, the FWHR strongly cued judgements of threat, and particularly judgements of aggressiveness as opposed to other indices of threat (e.g., untrustworthiness, prejudice), especially in younger faces. Sensitivity to the FWHR may be enhanced in younger male faces because young men have higher rates of violence and aggression than do the other demographic groups (Archer, 2009; M. Wilson & Daly, 1985). Likewise, the small but significant relationship between the larger FWHRs and dominance mirrored those for threat behaviour and were largely driven by men. The FWHR also cued dominance, but only for judgements of male faces. Although samples are predominantly restricted to studies among men, the small positive correlations between the FWHR and measures of athletic performance and success in business we found suggest a role of intra-sexual competition in shaping sex differences in the FWHR. In the dominance literature, several researchers have distinguished between social and physical forms of dominance (e.g., Puts, Gaulin, & Verdolini, 2006; Watkins, Jones, & DeBruine, 2010): individuals high in social dominance tend to be influential, respected, and a leader, whereas individuals high in physical dominance tend to be more capable of winning physical fights or contests against other members of the same sex. Most measures used in our meta-analysis of dominance involved questionnaires that captured better the construct of social rather than physical dominance, precluding our ability to formally test whether the type of dominance moderated the relationship. Similarly, in our analysis on the relationship between the FWHR and perceptions of dominance, only two of the included studies explicitly tapped into social dominance (the studies involved judgements of leadership and of social power, see supplementary tables), whereas other studies did not specify the type of dominance judgements they obtained. Further, one of the two studies produced an outlying effect size that was removed from the final analysis, again precluding our ability to test the type of dominance as a moderator. Future studies investigating the link between FWHR and perceived or actual dominance may benefit from including measures of both social and physical forms or judgements of dominance (e.g., see Hehman et al., 2014; Mileva, Cowan, Cobey, Knowles, & Little, 2014).

The FWHR is not the sole facial metric associated with masculine dominance and aggressiveness; studies have implicated jaw width (Cunningham, Barbee, & Pike, 1990), brow height, eye length, and mouth width (Toscano, Schubert, & Sell, 2014). However, the FWHR is well-situated in the upper face, where humans preferentially extract information about threat (Bassili, 1979). Further, whereas dominance and aggressiveness ratings of features such as jawlines may become enhanced by facial hair (Dixson &

Vasey, 2012), the association between ratings of aggressiveness and the FWHR is independent of beardedness (Geniole & McCormick, 2015). Perception of the FWHR involves low spatial frequency processing (Carré et al., 2010), as do social judgements (Todorov, Loehr, & Oosterhof, 2010), and thus the FWHR is perceived over longer distances than are specific facial features that rely on high spatial frequency processing. Low spatial frequency processing is rapid, as is assessment of the FWHR (Carré et al., 2010). Whether the FHWR cues judgements of threat and dominance because it subtly resembles angry expressions (e.g., the overgeneralization of emotional expression hypothesis, Said, Sebe, & Todorov, 2009) or because the emotional expression of anger simulates social cues of dominance and threat (Marsh, Adams, & Kleck, 2005) that become amplified by the FWHR remains to be determined. Nevertheless, our meta-analyses provide a starting point for addressing the hypothesis that the FWHR is part of an evolved cueing system of intra-sexual threat, dominance, and aggressiveness in men.

Tables

Table 5-1.

Summary of the final results of the meta-analyses conducted in the current manuscript.

Analysis	k	Mean Weighted Effect Sizes	95% Confidence Intervals	
			Low	High
Sex differences in the FWHR	32	$\bar{d}$ **= 0.11**	0.03	0.20
FWHR and Perceptions of Masculinity	12	$\bar{r}$ **= .30**	.18	.42
Stimuli sets of male faces	10	$\bar{r}$ **= .35**	.23	.47
Stimuli sets of female faces	2	$\bar{r}$ = -.01	-.26	.25
FWHR and Threat and Dominance Behaviour				
Threat Behaviour	31	$\bar{r}$ **= .13**	.09	.17
Within men	22	$\bar{r}$ **= .16**	.11	.21
Within women	9	$\bar{r}$ = .04	-.04	.12
Within Samples from North America	9	$\bar{r}$ **= .25**	.17	.33
Within Samples from Other areas	13	$\bar{r}$ **= .12**	.07	.17
Dominance Behaviour	16	$\bar{r}$ **= .12**	.05	.18
Success in Business-Related Outcomes	5	$\bar{r}$ **= .32**	.12	.50
Sports Performance	4	$\bar{r}$ **= .15**	.08	.22
FWHR and Perceptions of Threat and Dominance				
Perceptions of Threat				
Studies using a correlational design and/or a continua of faces with unmanipulated FWHRs	37	$\bar{r}$ **= .46**	.39	.53
Judgements were more strongly linked to the FWHR when faces were of younger than older individuals (k = 26, B = .40, p < .0001)				
Judgements of aggression were more strongly linked to the FWHR than were other types of judgements of threat (k = 26, B = .15, p = .01)				
Studies using manipulated FWHRs	7	$\bar{d}$ **= 0.41**	0.29	0.53
Perceptions of Dominance	7	$\bar{r}$ **= .20**	.06	.34
Stimuli sets of male faces only	4	$\bar{r}$ **= .30**	.19	.40
Stimulus sets of or including female faces	3	$\bar{r}$ = .06	-.19	.31
FWHR and Perceptions of Attractiveness	14	$\bar{r}$ **= -.26**	-.40	-.10
The negative relationship between judgements of attractiveness and the FWHR was stronger among samples with a greater than a lesser proportion of female observers (k = 14, B = .008, p = .01)				
FWHR and BMI	22	$\bar{r}$ **= .31**	.26	.36

Bolded effect sizes are significant and have confidence intervals that do not overlap zero ($p < .05$). k = number of samples included in the analysis. $\bar{d}$ = standardized mean difference, adjusted for small sample size bias. $\bar{r}$ = untransformed effect size coefficient (Pearson product moment correlation).

Figures

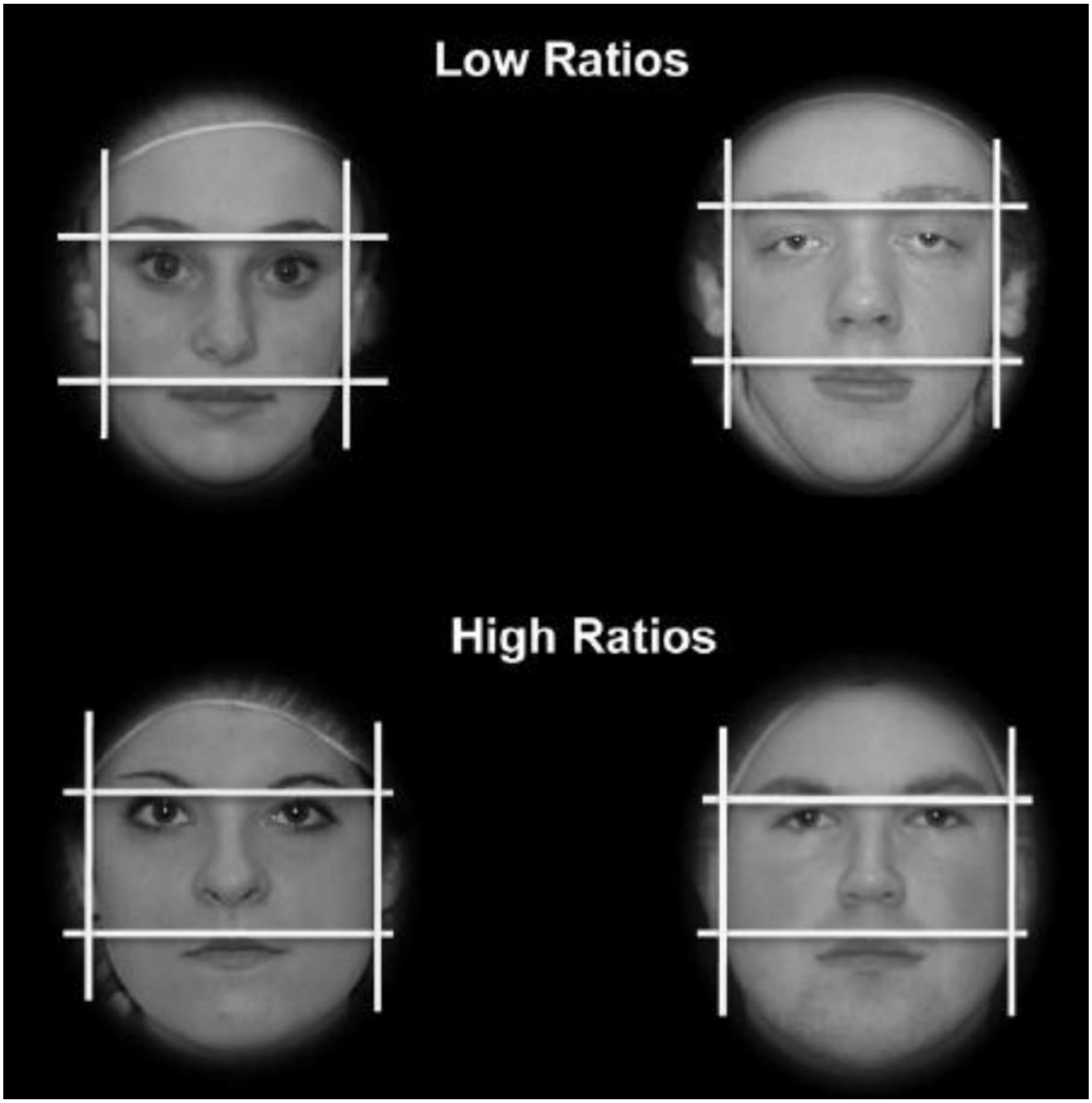

Figure 5-1. Examples of measurement of the FWHR in faces with relatively low and high FWHRs.

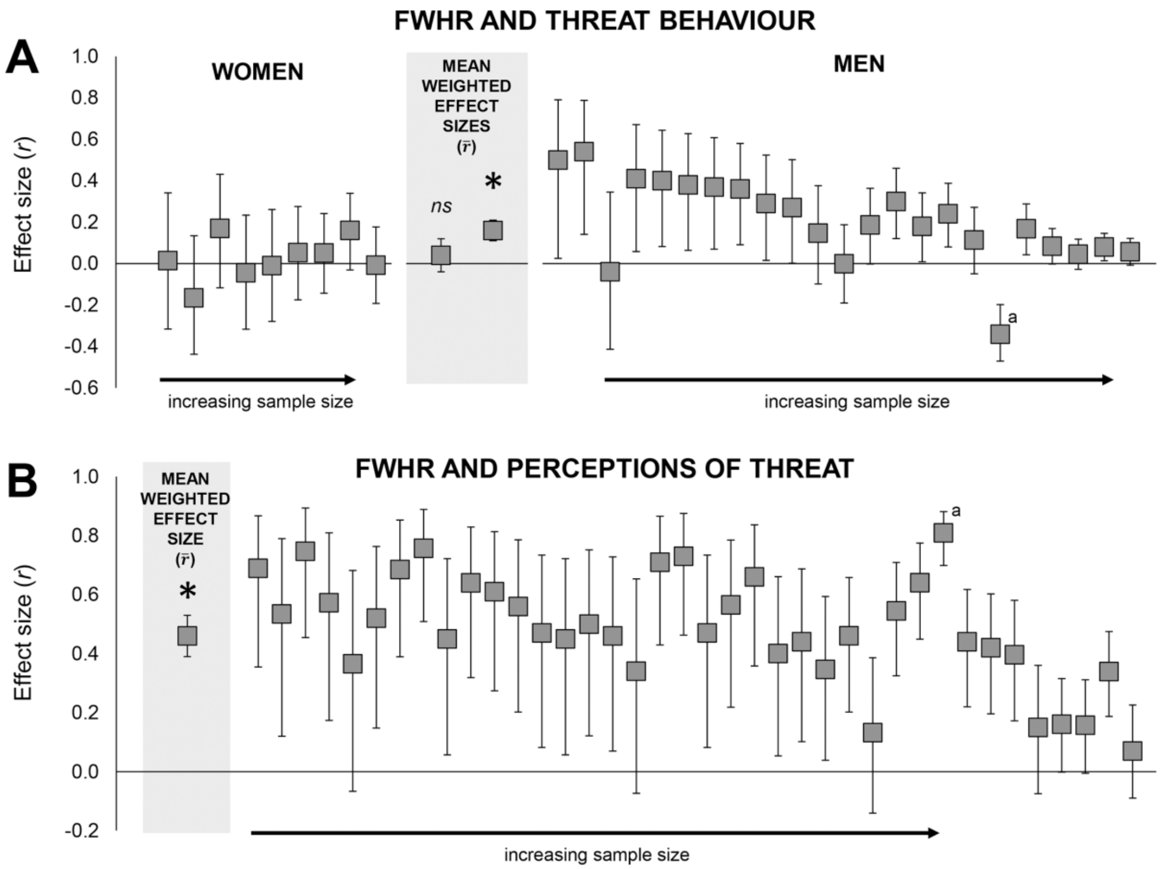

Figure 5-2. Effect sizes (*r*s) included in the meta-analysis on the relationships between the FWHR and threat behaviour (Panel A) and between the FWHR and perceptions of threat (Panel B). The mean weighted effect sizes **($\bar{r}$s)** are highlighted in grey, with men and women separated for Panel A and combined for Panel B. *$p < .0001$. [a]effect size was removed from the final analysis.

References

Abel, E. L., Kruger, M. L., & Dai, J. (2013). Changes in Male Testosterone Levels over the Last Century As Reflected In Facial Width-Height Ratios. *Journal of Andrology & Gynaecology*, *1*(1), 1–3.

Archer, J. (2009). Does sexual selection explain human sex differences in aggression? *Behavioral and Brain Sciences*, *32*(3-4), 249–66; discussion 266–311. http://doi.org/10.1017/S0140525X09990951

Arnott, G., & Elwood, R. W. (2009). Assessment of fighting ability in animal contests. *Animal Behaviour*, *77*(5), 991–1004. http://doi.org/10.1016/j.anbehav.2009.02.010

Bashir, N. Y., & Rule, N. O. (2014). Shopping under the influence: nonverbal appearance-based communicator cues affect consumer judgments. *Psychology of Marketing*, *31*, 539–548. http://doi.org/10.1002/mar

Bassili, J. N. (1979). Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology*, *37*(11), 2049–58. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/521902

Berry, D. S. (1991). Accuracy in Social Perception: Contributions of Facial and Vocal Information. *Journal of Personality and Social Psychology*, *61*(2), 298–307.

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*(4), 991–8. http://doi.org/10.1016/j.neubiorev.2010.10.016

Bonett, D. G. (2007). Transforming odds ratios into correlations for meta-analytic research. *The American Psychologist*, *62*(3), 254–5. http://doi.org/10.1037/0003-

066X.62.3.254

Bruce, V., & Young, A. (2012). *Face Perception*. New York, NY: Psychology Press.

Bryan, R., Perona, P., & Adolphs, R. (2012). Perspective distortion from interpersonal distance is an implicit visual cue for social judgments of faces. *PLoS ONE*, *7*(9), e45301. http://doi.org/10.1371/journal.pone.0045301

Carré, J. M. (2014). Social status, facial structure, and assertiveness in brown capuchin monkeys. *Frontiers in Psychology*, *5*(June), 567. http://doi.org/10.3389/fpsyg.2014.00567

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 of the Royal Society B Biological Sciences*, *275*(1651), 2651–6. http://doi.org/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*(10), 1194–8. http://doi.org/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. http://doi.org/10.1068/p6543

Cheng, J. T., Tracy, J. L., & Henrich, J. (2010). Pride, personality, and the evolutionary foundations of human social status. *Evolution and Human Behavior*, *31*, 334–347. http://doi.org/10.1016/j.evolhumbehav.2010.02.004.

Coetzee, V., Chen, J., Perrett, D. I., & Stephen, I. D. (2010). Deciphering faces: quantifiable visual cues to weight. *Perception, 39*(1), 51–61.

http://doi.org/10.1068/p6560

Cunningham, M. R., Barbee, A. P., & Pike, C. L. (1990). What do women want? Facialmetric assessment of multiple motives in the perception of male facial physical attractiveness. *Journal of Personality and Social Psychology*, *59*(1), 61–72. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2213490

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*(2), 235–238. http://doi.org/10.1016/j.jrp.2012.01.005

Dixson, B. J., & Vasey, P. L. (2012). Beards augment perceptions of men's age, social status, and aggressiveness, but not attractiveness. *Behavioral Ecology*, *23*(3), 481–490. http://doi.org/10.1093/beheco/arr214

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629–634. http://doi.org/10.1136/bmj.316.7129.469

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*(5), 1027–43. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24192545

Geniole, S. N., & McCormick, C. M. (2015). Facing our ancestors: Judgements of aggression are consistent and related to the facial width-to-height ratio in men irrespective of beards. *Evolution and Human Behavior*, *36*(4), 279–285. http://doi.org/10.1016/j.evolhumbehav.2014.12.005

Goetz, S. M. 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*(11), 2329–34. http://doi.org/10.1177/0956797613493294

Goldberg, L. R., Johnson, J. a., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*, 84–96. http://doi.org/10.1016/j.jrp.2005.08.007

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*(1), e52317. http://doi.org/10.1371/journal.pone.0052317

Grueter, C. C., Isler, K., & Dixson, B. J. (2015). Are badges of status adaptive in large complex primate groups? *Evolution and Human Behavior*. http://doi.org/10.1016/j.evolhumbehav.2015.03.003

Haselhuhn, M. P., Ormiston, M. E., & Wong, E. M. (2015). Men's facial width-to-height ratio predicts aggression: a meta-analysis. *PLOS ONE*, *10*(4), e0122637. http://doi.org/10.1371/journal.pone.0122637

Haselhuhn, M. P., Wong, E. M., Ormiston, M. E., Inesi, M. E., & Galinsky, A. D. (2014). Negotiating face-to-face: Men's facial structure predicts negotiation performance. *Leadership Quarterly*, *25*(5), 835–845. http://doi.org/10.1016/j.leaqua.2013.12.003

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hehman, E., Leitner, J. B., & Freeman, J. B. (2014). The Face – Time Continuum : Lifespan Changes in Facial Width-to-Height Ratio Impact Aging-Associated Perceptions. *Personality and Social Psychology Bulletin*, *40*, 1624–1636. http://doi.org/10.1177/0146167214552791

Kleiman, S., & Rule, N. O. (2013). Detecting suicidality from facial appearance. *Social Psychological and Personality Science*, *4*(4), 453–460. http://doi.org/10.1177/1948550612466115

Kramer, R. S. S., Jones, A. L., & Ward, R. (2012). A lack of sexual dimorphism in width-to-height ratio in white European faces using 2D photographs, 3D scans, and anthropometry. *PloS One*, *7*(8), e42705. http://doi.org/10.1371/journal.pone.0042705

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 signaling personality. *Evolution and Human Behavior*, *32*(3), 179–185. http://doi.org/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. *Evolutionary Psychology*, *10*(2), 320–337.

Lefevre, C. E., Lewis, G. J., Bates, T. C., Dzhelyova, M., Coetzee, V., Deary, I. J., & Perrett, D. I. (2012). No evidence for sexual dimorphism of facial width-to-height ratio in four large adult samples. *Evolution and Human Behavior*, *33*(6), 623–627. http://doi.org/10.1016/j.evolhumbehav.2012.03.002

Lefevre, C. E., Lewis, G. J., Perrett, D. I., & Penke, L. (2013). Telling facial metrics: facial width is associated with testosterone levels in men. *Evolution and Human*

Behavior, *34*(4), 273–279. http://doi.org/10.1016/j.evolhumbehav.2013.03.005

Lefevre, C. E., Wilson, V. a D., Morton, F. B., Brosnan, S. F., Paukner, A., & Bates, T. C. (2014). Facial width-to-height ratio relates to alpha status and assertive personality in capuchin monkeys. *PloS One*, *9*(4), e93369. http://doi.org/10.1371/journal.pone.0093369

Levine, S. Z. (2011). Elaboration on the association between IQ and parental SES with subsequent crime. *Personality and Individual Differences*, *50*(8), 1233–1237. http://doi.org/10.1016/j.paid.2011.02.016

Lewis, G. J. J., Lefevre, C. E., & Bates, T. C. C. (2012). Facial width-to-height ratio predicts achievement drive in US presidents. *Personality and Individual Differences*, *52*(7), 855–857. http://doi.org/10.1016/j.paid.2011.12.030

Lipsey, M. W., & Wilson, D. B. (2001). *Practical Meta-Analysis*. Thousand Oaks, CA: SAGE Publications, Inc.

Loehr, J., & Hara, R. B. O. (2013). Facial morphology predicts male fitness and rank but not survival in Second World War Finnish soldiers Facial morphology predicts male fitness and rank but not survival in Second World War Finnish soldiers, (May).

Marsh, A. A., Adams, R. B., & Kleck, R. E. (2005). Why do fear and anger look the way they do? Form and social function in facial expressions. *Personality & Social Psychology Bulletin*, *31*(1), 73–86. http://doi.org/10.1177/0146167204271306

Marsh, A. A., Cardinale, E. M., Chentsova-Dutton, Y. E., Grossman, M. R., & Krumpos, K. A. (2014). Power plays: Expressive mimicry of valid agonistic cues. *Social Psychological and Personality Science*, *5*(6), 684–690. http://doi.org/10.1177/1948550613519684

Mayew, W. J. (2013). Reassessing the association between facial structure and baseball performance: a comment on Tsujimura & Banissy (2013). *Biology Letters*, *9*, 20130538. http://doi.org/10.1098/rsbl.2013.0538

Maynard Smith, J., & Harper, D. (2003). *Animal signals*. New York, NY: Oxford University Press.

McGugin, R., & Gauthier, I. (2013). Face Recognition. In K. Ochsner & S. M. Kosslyn (Eds.), *The Oxford Handbook of Cognitive Neuroscience, Volume 2: The Cutting Edges* (p. 165). Oxford University Press.

Mileva, V. R., Cowan, M. L., Cobey, K. D., Knowles, K. K., & Little, a. C. (2014). In the face of dominance: Self-perceived and other-perceived dominance are positively associated with facial-width-to-height ratio in men. *Personality and Individual Differences*, *69*, 115–118. http://doi.org/10.1016/j.paid.2014.05.019

Orwin, R. G. (1983). A Fail-Safe N for effect size in meta-analysis. *Journal of Educational and Behavioral Statistics*, *8*(2), 157–159. http://doi.org/10.3102/10769986008002157

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*(3), 169–173. http://doi.org/10.1016/j.evolhumbehav.2011.08.001

Parker, G. A. (1974). Assessment strategy and the evolution of fighting behaviour. *Journal of Theoretical Biology*, *47*(1), 223–243. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4477626

Peterson, R. a, & Brown, S. P. (2005). On the use of beta coefficients in meta-analysis. *The Journal of Applied Psychology*, *90*(1), 175–181. http://doi.org/10.1037/0021-

9010.90.1.175

Puts, D. A. (2010). Beauty and the beast: mechanisms of sexual selection in humans. *Evolution and Human Behavior*, *31*(3), 157–175. http://doi.org/10.1016/j.evolhumbehav.2010.02.005

Puts, D. A., Gaulin, S. J. C., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior*, *27*(4), 283–296. http://doi.org/10.1016/j.evolhumbehav.2005.11.003

Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion (Washington, D.C.)*, *9*(2), 260–4. http://doi.org/10.1037/a0014681

Sanchez-pages, S., Rodriguez-ruiz, C., & Turiegano, E. (2014). Facial Masculinity : How the Choice of Measurement Method Enables to Detect Its Influence on Behaviour. *PLOS ONE*, *9*(11), e112157. http://doi.org/10.1371/journal.pone.0112157

Schulz, M. S., Cowan, C. P., & Cowan, P. a. (2006). Promoting healthy beginnings: a randomized controlled trial of a preventive intervention to preserve marital quality during the transition to parenthood. *Journal of Consulting and Clinical Psychology*, *74*(1), 20–31. http://doi.org/10.1037/0022-006X.74.1.20

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 of the Royal Society B: Biological Sciences*, *276*(1656), 575–84. http://doi.org/10.1098/rspb.2008.1177

Sterne, J. A. C., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology*, *54*(10), 1046–1055.

http://doi.org/10.1016/S0895-4356(01)00377-8

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*(4), 298–303. http://doi.org/10.1016/j.evolhumbehav.2009.12.001

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*(5), 551–556. http://doi.org/10.1016/j.evolhumbehav.2012.02.002

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics* (5th ed.). Boston, MA: Pearson.

Tibbetts, E. A., & Lindsay, R. (2008). Visual signals of status and rival assessment in Polistes dominulus paper wasps. *Biology Letters*, *4*(3), 237–9. http://doi.org/10.1098/rsbl.2008.0048

Todorov, A., Loehr, V., & Oosterhof, N. N. (2010). The obligatory nature of holistic processing of faces in social judgments. *Perception*, *39*, 514–532. http://doi.org/10.1068/p6501

Toscano, H., Schubert, T. W., & Sell, A. N. (2014). Judgments of dominance from the face track physical strength. *Evolutionary Psychology*, *12*(1), 1–18.

Třebický, V., Fialová, J., Kleisner, K., Roberts, S. C., Little, A. C., & Havlíček, J. (2015). Further Evidence for Links Between Facial Width-to-Height Ratio and Fighting Success : Commentary on Zilioli et al.(2014). *Aggressive Behavior*, *41*, 331–334. http://doi.org/10.1002/ab.21559

Wang, Y., Geigel, J., & Herbert, A. (2013). Reading personality: Avatar vs. human faces. *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*, 479–484. http://doi.org/10.1109/ACII.2013.85

Watkins, C. D., Jones, B. C., & DeBruine, L. M. (2010). Individual differences in dominance perception: Dominant men are less sensitive to facial cues of male dominance. *Personality and Individual Differences*, *49*, 967–971. http://doi.org/10.1016/j.paid.2010.08.006

Welker, K. M., Goetz, S. M. M., & Carré, J. M. (2015). Perceived and experimentally manipulated status moderate the relationship between facial structure and risk-taking. *Evolution and Human Behavior*. http://doi.org/10.1016/j.evolhumbehav.2015.03.006

Welker, K. M., Goetz, S. M. M., Galicia, S., Liphardt, J., & Carré, J. M. (2014). An Examination of the Associations Between Facial Structure, Aggressive Behavior, and Performance in the 2010 World Cup Association Football Players. *Adaptive Human Behavior and Physiology*. http://doi.org/10.1007/s40750-014-0003-3

Weston, E. M., Friday, A. E., & Liò, P. (2007). Biometric evidence that sexual selection has shaped the hominin face. *PloS One*, *2*(8), e710. http://doi.org/10.1371/journal.pone.0000710

Wilson, D. B. (2001). Effect size determination program. College Park, MD: University of Maryland.

Wilson, M., & Daly, M. (1985). Competitiveness, risk taking, and violence: the young male syndrome. *Ethology and Sociobiology*, *6*(1), 59–73. http://doi.org/10.1016/0162-3095(85)90041-X

Wilson, V., Lefevre, C. E., Morton, F. B., Brosnan, S. F., 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. http://doi.org/10.1016/j.paid.2013.10.008

Zilioli, S., Sell, A. N., Stirrat, M., Jagore, J., Vickerman, W., & Watson, N. V. (2015). Face of a fighter: Bizygomatic width as a cue of formidability. *Aggressive Behavior*, *41*, 322–330. http://doi.org/10.1002/ab.21544

Chapter 6: General Discussion and Conclusions

Summary of thesis findings and their contribution to the field of social perception

Across these chapters, I have shown that: the face ratio is associated with objectively measured cheating behaviour, likely because it is correlated with personality traits related to psychopathy (fearless dominance) that influence one's willingness to commit such behaviours, providing ecological validity for the metric as an advertisement of one's personality and threat potential (Chapter 2); the face ratio shares stronger and more direct links with judgements of aggression than with judgements of trustworthiness, which led me to conclude initially that the face ratio may be primarily a signal of aggressiveness (Chapter 3)[20]; the face ratio maintains its association with judgements of aggression even when men were bearded, supporting the idea that it may have been operational in our ancestral past, and (Chapter 4); that when effect sizes are combined across numerous studies, involving participants from around the world (e.g., China, Japan, UK, Canada, USA, Mexico, Germany, Czech Republic), the face ratio shares significant (albeit relatively weak) associations with actual threat behaviour, and strongly cues judgements of threat potential (Chapter 5). Together, this body of work provides some initial evidence that the face ratio, and sensitivity to it, may be part of an evolved system designed for advertising and assessing threat potential, similar to other facial

[20]Based on the strong correlations between judgements of trustworthiness and judgements of aggression (see Chapter 3), and the face ratio's associations with both of these judgements, however, the face ratio may be best characterized as cuing threat potential more generally, and that stronger links between the face ratio and judgements of aggression versus trustworthiness may exist simply because judgements of aggression are perhaps a more accessible and representative indicator of this construct than are judgements of trustworthiness. Consistent with this idea, facial judgements appear to load onto two dimensions, trustworthiness and dominance, with threat involving a combination of low trust and high dominance (Oosterhof & Todorov, 2008). Aggression appears to represent a similar combination of these factors (see Table S3 in Oosterhof & Todorov, 2008) and may thus best approximate this larger threat potential construct than trustworthiness (or dominance) individually.

mechanisms proposed to communicate threat potential and dominance in other species [e.g., darkness of facemask in male eland antelopes (*Tragelaphus oryx*), Bro-Jørgensen & Beeston, 2015; facial redness in male mandrills (*Mandrillus sphinx*), Setchell et al., 2008; black facial patterns in paper wasps (*Polistes dominulus*), Tibbetts & Dale, 2004; Tibbetts & Lindsay, 2008]. Therefore, the face ratio may represent one mechanism through which observers in previous studies were able to accurately assess, from photographs of faces, threat-related traits such as power (Berry, 1991; see also Berry, 1990), Machiavellianism and other dark triad traits (Cherulnik et al., 1981; Holtzman, 2011), aggressiveness (Trebicky, Havlícek, Roberts, Little, & Kleisner, 2013), dominance (Mueller & Mazur, 1996; this judgement also tracked upper body strength in some studies, see Fink et al., 2007; Toscano et al., 2014; Undurraga et al., 2010), propensity for violence (Stillman et al., 2010), and strength (Holzleitner & Perrett, 2016; Sell et al., 2009).

Research from other labs has also shown that the relative size of the metric is associated with dominance style across species of macaques (*Macaca*) (Borgi & Majolo, 2016) and with alpha status and assertiveness (Lefevre, Wilson, et al., 2014; V. Wilson et al., 2014) in capuchin monkeys (*Sapajus spp*), with the link between the face ratio and assertiveness particularly pronounced among capuchins low in status (Carré, 2014). Notably, the shape of this interaction parallels the shape of the interaction between the face ratio and status in predicting aggression and risk-taking in humans (Goetz et al., 2013; Welker, Goetz, & Carré, 2015b), an interaction that may explain some of the heterogeneity and overall weakness of the effect size (see below, Why are observers so sensitive to the face ratio if it possesses limited ecological validity?). Although no studies have investigated whether non-human primates utilize this metric when making fight/flight or approach/avoidance decisions, there is evidence that humans form reliable and accurate judgements of dominance from the faces of chimpanzees (*Pan troglodytes*) (Kramer, King, & Ward, 2011; Kramer & Ward, 2012), providing some initial support for a shared advertisement and assessment system across human and non-human primates. Children and adults also appear sensitive to this metric when judging the aggressiveness of same- and other-race faces, suggesting that this sensitivity does not depend on experience with a given face category (Short et al., 2012). Together with my thesis work, these studies suggest that the face ratio and sensitivity to it may be part of an evolved advertisement and assessment system of threat potential.

It will be important to examine how early in development observers begin to show sensitivity to the face ratio. Threat detection emerges spontaneously without social experience in rhesus monkeys (*Macaca mulatta*) (Sackett, 1966). Threat detection is also evident early in development in humans (reviewed in LoBue & Rakison, 2013), with preverbal infants (~10 month-olds) using markers of threat potential (relative size) to predict contest outcomes between rival agents (Thomsen, Frankenhuis, Ingold-Smith, & Carey, 2011). The extent to which such young infants utilize the face ratio is unknown but this question is an important avenue for future research.

In addition to providing evidence for the ecological validity of this metric, I also examined a correlate of the face ratio that may be more directly linked to behaviour: personality traits related to psychopathy (Chapter 2). My finding that these personality traits were more directly linked to behaviour than the face ratio and, when controlled statistically, reduced the association between the face ratio and behaviour, hints at a correlate of the face ratio that more directly determines threat potential (see also

Haselhuhn & Wong, 2012). Further, the association between the face ratio and these psychopathic personality traits (fearless dominance), specifically, has been replicated in a sample involving undergrads and prison inmates (although the relationship appears to have been driven by the prison inmates, Anderl et al., 2016), and my meta-analysis also revealed a significant (but weak) association between the face ratio and dominance as measured by a variety of other relevant indices (Chapter 5). Therefore, the face ratio may be linked with threat behaviour because it is correlated with dominance, a trait relevant to threat potential. Nevertheless, these findings do not explain how such a relationship between the face and personality may emerge. Below, I provide a review of some potential mechanisms that may explain links between the face ratio, personality, and behaviour.

Another unanswered question from my thesis studies is why are the magnitudes of the associations between the face ratio and behaviour, and between the face ratio and judgements, so discrepant? In other words, why are observers so sensitive to the face ratio if it possesses rather weak ecological validity? Additionally, irrespective of the ecological validity of this metric, what are the social consequences of these judgements that form based on the face ratio? If perceived as more threatening, are men with larger face ratios also treated more poorly in social interactions? Do people intentionally alter their face ratios to achieve certain goals in interpersonal settings? Further, how did the sex difference (albeit small) in the face ratio evolve and why are observers sensitive to the face ratio in women if, across studies, it often shares non-significant and smaller links with behaviour than it does in men? Below, I address each of these questions, in turn.

What mechanisms may explain links between the face ratio, psychological traits, and behavioural tendencies?

According to the Developmental Model of Relationships Between Physical and Psychological Qualities (Zebrowitz & Collins, 1997), there are multiple routes through which physical qualities can be associated with psychological qualities (and thus behaviour). One explanation is that variation in both variables is caused by a third, common biological factor (e.g., consider the biological factors that influence both the mental abilities and the physical features of individuals with Down's Syndrome and Fetal Alcohol Syndrome, Zebrowitz & Collins, 1997). Through the Advertisement, Assessment, and Action model described in the General Introduction (Chapter 1), I proposed a similar underlying, biological mechanism. One biological factor that has been proposed to account for associations between the face ratio, personality traits, and threat behaviour, is the effects of pubertal testosterone (Carré & McCormick, 2008). Surges in pubertal testosterone influence face shape (e.g., Marečková et al., 2011; Verdonck, Gaethofs, Carels, & de Zegher, 1999) and also neural circuitry involved in social behaviour (reviewed in Blakemore, 2012; Schulz, Molenda-Figueira, & Sisk, 2009), organizational changes that can have long-lasting consequences. In one study, for example, pubertal testosterone concentrations predicted the occurrence of problem behaviour years later (Drigotas & Udry, 1993).

One recent study (Hodges-Simeon, Hanson Sobraske, Samore, Gurven, & Gaulin, 2016) examined the relationship between the face ratio and concentrations of testosterone using a cross-sectional sample of males, aged 8 - 23 years. The authors reported a significant association between the face ratio and testosterone concentrations when controlling statistically for age. It will be important to extend these findings and determine, using longitudinal datasets, whether hormones in adolescence predict the size of the face ratio in adulthood. Other researchers have identified relationships between the face ratio and concentrations of testosterone in adulthood (Lefevre, Lewis, Perrett, & Penke, 2013). These associations were not replicated in subsequent, lager studies, however (Bird et al., 2016). It is also possible that prenatal biological mechanisms explain links between the face ratio and threat potential in adulthood. For example, Zebrowitz, Franklin, and Boshyan (2015) found that infants who were lower in distress had larger face ratios at 14 months of age than did infants who were higher in distress. This study provided some initial evidence that the link between the face ratio and behaviour may emerge earlier than would be expected based on hypotheses involving pubertal hormones; prenatal steroid hormones may influence both face shape and temperament and explain such early links between these variables, associations that may persist into adulthood (Zebrowitz et al., 2015). One study examined testosterone concentrations in the blood of umbilical cords and the corresponding face shape of the participants 20 years later (Whitehouse et al., 2015). The authors reported a small marginal association between these prenatal testosterone concentrations and the adult face ratio in men, but not women. Nevertheless, other putative indices of prenatal testosterone tend to share weak associations with threat potential in adulthood (meta-analysis in Hönekopp & Watson, 2011), reducing the likelihood that this factor represents a common variable explaining the associations between the face ratio and behaviour in adulthood.

Another mechanism posited to underlie links between physical and psychological traits is self-fulfilling prophesy (Zebrowitz & Collins, 1997). Rather than men with larger face ratios being more predisposed to aggressive and antisocial behaviour, they may instead be expected by observers to behave in such ways and thus, as a consequence of these expectations, be treated differently. This differential treatment, in turn, may elicit the anticipated negative behaviours. For example, one study found that participants preferred to share less of their resources with men with larger (vs smaller) face ratios because they believed such men would act selfishly themselves (Haselhuhn, Wong, & Ormiston, 2013). The authors then investigated whether being treated like a man with a larger (vs smaller) face ratio would alter subsequent selfish behaviour. They found that participants who were treated like men with larger face ratios by their bargaining partner (i.e., received less fair shares of the resource) acted more selfishly towards their bargaining partner in a subsequent task than did participants who were treated like men with smaller face ratios (i.e., received fairer shares of the resource). Therefore, men with larger face ratios may be treated differently than men with smaller face ratios and this differential treatment, in turn, may elicit more threatening behaviours in high face ratio than in low face ratio men (for evidence of similar self-fulfilling prophecy effects regarding facial structure and honesty, see Zebrowitz, Voinescu, & Collins, 1996). One limitation to this proposed mechanism, however, is that it does not explain how the perception of men with larger face ratios as more threatening initially began.

Environmental and psychological factors can also influence craniofacial shape (reviewed in Zebrowitz & Collins, 1997). For example, different diets (e.g., tougher and less processed foods versus softer and more processed foods) require more or less chewing-related muscle activity (e.g., Paschetta et al., 2010) and environmental and psychological stressors can lead some individuals to habitually clench their jaw (as in bruxism, e.g., Serra-Negra, Paiva, Flores-Mendoza, Ramos-Jorge, & Pordeus, 2012). Such changes in muscle activity can influence craniofacial growth (e.g., Ingervall & Bitsanis, 1987, reviewed in Kiliaridis, 1995). In some cases, the same traits (e.g., neuroticism) that lead to the habitual clenching of the jaw (e.g., Sutin, Terracciano, Ferrucci, & Costa, 2010) may also promote antisocial and aggressive behaviours (e.g., Castellani et al., 2014; meta-analysis in Jones, Miller, & Lynam, 2011) and thoughts (e.g., Boduszek, Shevlin, Adamson, & Hyland, 2013). Therefore, common psychological traits may influence both the shape of an individual's face and their behavioural propensities. A more direct investigation of the common traits underlying habitual jaw clenching and the face ratio may be worthwhile in future studies.

Rather than through jaw- and chewing-related muscle activity, certain personality traits may also influence face structure through the regular use of emotional expressions (Zebrowitz & Collins, 1997). For example, older adults that were judged as more angry looking (based on photographs of their faces posed in neutral expressions), self-reported higher levels of trait anger than did older adults that were judged as less angry looking (Malatesta, Fiore, & Messina, 1987). The authors suggested that persistent mood states likely altered the appearance of the older-adult faces over time, and did so in a way that made the resting or neutral face subtly resemble the corresponding emotion that was regularly experienced. It is unlikely, however, that these slow changes explain associations between the face ratio and judgements of threat, which are often found in young adults and is related to variation in bone rather than skin structure/appearance.

Why are observers so sensitive to the face ratio if it possesses limited ecological validity?

Although the face ratio was only weakly linked to threat behaviours, observers appeared to strongly utilize this metric for judgements of threat. Why are observers much more sensitive to the face ratio than would be expected based on its association with real behaviour? One possibility is that the relatively weak association between the face ratio and behaviour identified in Chapter 5 is not an accurate reflection of the true relationship. Instead, this relationship may have been weak because of the high level of error variability associated with taking photographs of a face (Todorov & Porter, 2014; see also Jenkins, White, Van Montfort, & Mike Burton, 2011). Slight variations in the tilt of the participant’s head (Hehman, Leitner, & Gaertner, 2013) and the distance between the camera lens and the face (Bryan, Perona, & Adolphs, 2012; see also Kramer, 2016) may skew the face ratio and, possibly, social judgements of the individual (Todorov & Porter, 2014)[21]. In fact, that there was any relationship between the face ratio and behaviour over and above this “noise” makes the relationship, albeit small, more impressive. On the other hand, recall the measurements of the face ratio I reported from the non-bearded and the bearded versions of the same men’s faces in Chapter 4. These photographs were taken weeks to months apart and involved substantial changes to the appearance of the face (beard growth) yet there was still a strong correlation between the social judgements of the two sets of faces. Therefore, if care is taken to ensure the faces are forward facing with minimal tilt, it is possible to reduce such error variability associated with facial

[21]Although judgements based on static cues in the face (such as the face ratio) are more stable across different photos of the same individual than are judgements based on dynamic cues, such as those related to emotional expressions (Hehman, Flake, et al., 2015).

photographs. Further, physical measurements of the face ratio using calipers, which are less prone to such measurement error, share strong correlations with measures obtained from photographs (Kramer, Jones, & Ward, 2012), again suggesting that it is possible to minimize this error variability[22]. Much like measurements of faces can vary from photograph to photograph, measurements of behaviour can vary from situation to situation, adding additional error variability that may obscure accurate estimates of the true effect size. Therefore, to provide the best estimates of the association between the face ratio and behaviour, future studies might benefit from using multiple facial photographs and multiple behavioural measures of threat potential; using these aggregates would reduce any photo- and situation- specific measurement error and possibly yield larger effect sizes that may be more similar to those obtained between the face ratio and observer judgements of threat potential.

Another possibility is that the face ratio's ecological validity varies across types of individuals, but observers overgeneralize its validity to all populations. For example, across two studies, the relationship between the face ratio and aggression was moderated by the individual's socioeconomic status: the face ratio was associated with aggression among men low, but not high, in socioeconomic status (Goetz et al., 2013). A similar interaction between the face ratio and status emerged when investigating risk-taking (Welker, Goetz, & Carré, 2015a) (for similar interactions involving socioeconomic status and other face traits, see Zebrowitz, Andreoletti, Collins, Lee, & Blumenthal, 1998). Goetz and colleagues (2013) proposed that the association between the face ratio and

[22] Additionally, if such noise was a factor, we would also expect lower associations between the face ratio and social judgements than the ones I obtained throughout this thesis. In other words, such "noise" should reduce associations between both the face ratio and behaviour and between the face ratio and social judgements.

aggression may be enhanced among those low in status because such individuals have much less to lose (further losses in status are not possible) and more to gain (increases in status and reputation and thus access to valued social resources) from aggressive interactions than do those higher in status. Thus, the face ratio may possess strong ecological validity among low but not high status men, but observers overgeneralize their use of this metric to men with high status as well[23].

Another possibility is that the face ratio does possess weak ecological validity but because the costs required to maintain sensitivity to this metric and to mislabel someone as threatening is often lower than the potential costs associated with ignoring the cue and being exploited or harmed, it is still in the observer's best interest to attend to and utilize this metric in most situations. According to Error Management Theory (Haselton & Buss, 2000; also see Haselton & Funder, 2006), biases in perception should arise anytime there are or were asymmetries in the costs associated with making these two types of errors (incorrectly judging someone as threatening vs incorrectly judging someone as non-threatening): Selection pressures should favour cognitive mechanisms that make many cheap errors over those that make few fatal ones. Therefore, this asymmetry should, over the course of history, lead to a perceptual bias in which observers consistently judge men with high face ratios as more aggressive despite the metric possessing limited ecological validity.

Other researchers have proposed that social judgements based on neutral faces depend on the face's subtle resemblance to a given emotional expression (e.g., Franklin

[23]One study found, however, a non-significant association between judgements of aggression and the face ratio of businessmen who are, presumably, of higher socioeconomic status than are other members of the population (Alrajih & Ward, 2014). Therefore, it is also possible that observers do utilize the face ratio differently for low versus for high status individuals.

& Zebrowitz, 2013; Montepare & Dobish, 2003; Said, Sebe, & Todorov, 2009). As pointed out by Carré and colleagues (2009), angry expressions, which involve the lowering of the brow and the raising of the upper lip, increase the size of the face ratio. Observers may thus perceive men with larger face ratios as more aggressive simply because they appear angrier than men with smaller face ratios. In contrast, Marsh and colleagues (2005, 2014) have proposed that the relationship between angry expressions and threatening facial features is reversed. Rather than static features looking more or less threatening based on their resemblance to an angry expression, the authors propose that the angry expression evolved to look the way it does because this pattern of facial movements increases the size of the face ratio and thus mimics the morphological features of a more threatening face. Indeed, angry expressions increased the perceived aggressiveness of a face to the extent that they increased the size of the face ratio (Marsh et al., 2014); in other words, changes in the size of the face ratio mediated partially the influence of angry expressions on judgements of aggression. Therefore, angry expressions may be effective at communicating threat because they make the face ratio appear larger. Nevertheless, regardless of the direction of this relationship between face morphology and emotional expressions, the face ratio maintains its association with judgements of aggression when controlling statistically for judgements of anger (Boshyan et al., 2014), and when controlling experimentally for the facial muscles involved in the expression of anger (Marsh et al., 2014). Therefore, the face ratio’s resemblance to an angry expression cannot fully account for observer utilization of this metric.

Although this thesis has focused on men’s ability and propensity to cause trouble or harm others, some of observer sensitivity to the face ratio may instead be explained by these men’s ability to defend themselves or withstand attack from others. For example, the majority of assault-related injuries are to the face, with the zygoma (the bone that accounts for the width of the face ratio) being one of the most common fracture sites (reviewed in Carrier & Morgan, 2014). Variation in the face ratio (and the sex difference in the size of this metric) may have thus evolved to buttress the male face from physical attacks (Carrier & Morgan, 2014; also see Puts, 2010). Men with larger face ratios are less likely to die from contact violence than are men with smaller face ratios (Stirrat, Stulp, & Pollet, 2012) lending some support to this buttressing hypothesis. Further, when participants are forced to select an opponent in a competitive interaction, they more often choose men with smaller rather than larger face ratios (Hehman, Leitner, Deegan, & Gaertner, 2015). Therefore, part of the association between men’s face ratios and judgements of threat potential might be driven by the fact that face ratios advertise an individual’s ability to withstand an attack rather than to initiate one. For this reason, people may actively avoid physical competitions against such individuals.

Observers may also be sensitive to the face ratio because it involves the relative spacing between or configuration of features (e.g., the distance between the left and right zygion divided by the distance between the mid-brow and the upper lip) rather than the size or shape of the individual features themselves. Configural information from the face is processed more rapidly, and appears to more strongly guide judgements of aggression and of threat, than does featural information (e.g., Bar et al., 2006; Carré et al., 2010). Further, the ratio of the width and height components is more effective at guiding social judgements than either component alone (Carré et al., 2010), likely because our visual system processes faces holistically (e.g., the width and height in combination, rather than either feature alone) (reviewed in Maurer, Le Grand, & Mondloch, 2002). Supplementary analyses of data in Chapter 3 (see Supplementary Analyses), however, indicate that when the width and height components are allowed to vary independently of one another within the same regression model, they account for approximately the same amount of variability in judgements of aggression as does the face ratio. Such results suggest that the face ratio is not a more (or less) effective predictor of aggression judgements than the separate width and height components when considered jointly. Nonetheless, from a practical perspective, researchers may prefer to use the face ratio alone rather than these two individual width and height components because doing so requires one less degree of freedom (thus increasing statistical power in regression models).

Conversely, the ability to examine both components simultaneously may provide additional insight into the perceptual information used by observers to make social judgements that is not offered when using the face ratio alone. For example, supplementary analyses of data from Chapter 4 (see Supplementary Information) revealed that although the height and width components are correlated positively (i.e., individuals with taller upper faces tend to also have wider faces than those with shorter upper faces), highlighting the existence of a more general "upper face size" factor, each component contributes uniquely and in opposite direction to the prediction of judgements of aggression: Holding the height of the face constant, individuals with wider faces appear more aggressive than those with narrower faces. To the extent that individuals with bigger upper faces also have bigger bodies (and thus may cause more damage in an agonistic interaction), this result might not be too surprising. On the other hand, the unique and opposing effects also indicate that, holding the width of the face constant, individuals with shorter upper faces appear more aggressive than do those with taller upper faces, which is opposite to what one might predict based on the overall size of the upper face. Therefore, by entering the two separate components as simultaneous predictors within the same model, we can better identify how the specific shape of the upper face, rather than its overall size, contributes to social judgements.

Investigating both components, rather than the ratio, may also provide insight into the sex specificity of the relationship between the face ratio and actual threat behaviour. One possibility is that the face ratio is correlated weakly with threat behaviour in women because, unlike in men, only one of the two components possesses ecological validity (i.e., is associated with behaviour). If so, using the face ratio rather than each component may have obscured relationships investigated previously in women.

What are the social consequences of the judgements that are based on the face ratio?

Although many studies have investigated the extent to which social judgements are associated with the face ratio, perhaps a more important question is what are the social consequences of these judgements in interpersonal interactions? Men with larger face ratios are perceived as more selfish and as less trustworthy and thus people are less generous and less trusting of such men when sharing and investing resources (e.g., Efferson & Vogt, 2013; Haselhuhn et al., 2013; Stirrat & Perrett, 2010). Further, men with larger (vs smaller) face ratios were more likely to be sentenced to death rather than to life in prison (J. P. Wilson & Rule, 2015), although it is possible that such individuals were more violent and thus deserving of this sentence.

On the other hand, the perceived aggressiveness and dominance associated with the face ratio may be beneficial in certain types of interpersonal interactions. For example, despite receiving lower shares of resources in the economic tasks mentioned above, such men may receive larger shares of a resource when there is opportunity for them to retaliate in response to any unfair treatment they receive. In the Ultimatum Game (designed by Güth, Schmittberger, & Schwarze, 1982), a participant is given a sum of money and is asked to share the sum of money with another individual. Although the participant can share as much or as little of the resource as they want, they are told that the other individual can reject offers that they deem unfair (retaliatory punishment), which results in both players receiving zero dollars. Participants must therefore weigh the benefits of keeping a larger share of money with the potential costs of retaliatory punishment (i.e., having their offer rejected). In this task, men with larger face ratios received larger offers than did men with smaller face ratios (MacDonell, Geniole, & McCormick, 2015), a relationship that was mediated by perceptions of such men as being more aggressive. Further, this relationship between the face ratio and offers was most pronounced when the offers were made by weaker (vs stronger) men. Therefore, in economic interactions in which unfair treatment can result in retaliation, men with larger face ratios are treated more fairly (i.e., receive larger economic offers, closer to 50% of the total resource and, in some cases, more than 50% of the total resources) than men with smaller face ratios, especially by others who are vulnerable to harm during physical conflict. As mentioned above, participants also avoid competing against men with larger face ratios, and instead prefer to choose such individuals as partners in physical and competitive activities (Hehman, Leitner, et al., 2015). Therefore, despite unfair treatment in some situations, the greater perceived threat potential among men with larger face ratios may enable them to more readily gain and maintain dominance in social hierarchies. Indeed, Chapter 5 revealed that men with larger face ratios are often perceived as more dominant and have greater success in business-related outcomes and in competitive sports.

There is also evidence that some individuals utilize this metric to their advantage in interpersonal interactions. For example, when asked to appear intimidating, strong, powerful, or prone to attack, people tilt their head (Hehman et al., 2013) or show an angry expression (Marsh et al., 2014), both of which increases the relative size of the face ratio. Therefore, not only do perceptions based on the face ratio influence behaviour in social interactions, but participants appear to intentionally alter this metric when trying to appear tough and intimidating.

How did the face ratio evolve to be larger in men than in women, and why are observers so sensitive to this metric in women if it is not associated with female threat behaviour?

There has been mixed evidence regarding the size of the face ratio in men and in women. Initial studies reported significant sex differences, with male face ratios larger than female face ratios (e.g., Carré & McCormick, 2008; Weston et al., 2007). Other labs, however, reported that there were no sex differences in this metric (e.g., Kramer et al., 2012; Lefevre et al., 2012) but have since provided evidence to the contrary (e.g., Kramer, 2015; Lefevre, Etchells, Howell, Clark, & Penton-voak, 2014). In my meta-analysis in Chapter 5, I conducted the most statistically powerful test of this sex difference to date, and found that men had significantly (albeit slightly) larger face ratios than did women. Therefore, there appears to be a small, but significant sex difference in the size of the face ratio.

Weston and colleagues (2007) suggested that intersexual selection pressures may drive this sex difference in the size of the face ratio. Further, two studies provided evidence that men with larger (vs smaller) face ratios have greater reproductive fitness, operationalized by number of children (Loehr & Hara, 2013) or number of children that reached an age of at least 15 years (Gómez-Valdés et al., 2013). Nevertheless, the mechanism driving the enhanced reproductive fitness among these high face ratio men is not clear. For example, my meta-analysis in Chapter 5 revealed that observers, especially female observers, perceive men with larger face ratios as less attractive than men with smaller face ratios. In another study, women perceived men with larger face ratios as less desirable for both short- and long- term relationships than men with smaller face ratios (Geniole & McCormick, 2013), thus raising doubts about this sex difference in the size of the face ratio arising from attractiveness-driven, intersexual selection pressures.

Alternatively, differences in the shape of the male and female face likely arose from competition-based intrasexual selection pressures (Puts, 2010). Across several studies, men with larger face ratios were better able to extract resources (see meta-analysis on success in business in Chapter 5; Haselhuhn, Wong, Ormiston, Inesi, & Galinsky, 2014; MacDonell et al., 2015; Wong, Ormiston, & Haselhuhn, 2011) and were higher in both actual and perceived dominance (see meta-analysis in Chapter 5; Alrajih & Ward, 2014; Lefevre, Etchells, et al., 2014; Valentine, Li, Penke, & Perrett, 2014) than were men with smaller face ratios. Further, women value resource acquisition abilities in a potential mate more than do men (e.g., Buss, 1989) and the value placed on this ability is greater than the value women place on the attractiveness of the potential mate (e.g., Li, Bailey, Kenrick, & Linsenmeier, 2002). Therefore, the face ratio may have become sexually dimorphic because men with larger face ratios had competitive advantages over men with smaller face ratios, and these competitive advantages (dominance), rather than the larger face ratio, per se, were desirable to women. Consistent with this idea, one study found that the association between the face ratio and women's interest in a short-term relationship was mediated partially by perceptions of such men being more dominant than men with smaller face ratios (Valentine et al., 2014). It will be important for future studies to more directly investigate links between the face ratio, resource acquisition abilities, and reproductive success.

Because women typically lack the threat potential to physically outcompete men in agonistic encounters (based on a number of indices of threat potential, reviewed in Sell, Hone, & Pound, 2012), the face ratio may have evolved to primarily be an advertisement and assessment system of threat for inter-male than for inter-female competitions. Consistent with this idea, the meta-analysis showed that the face ratio is an ecologically valid cue of threat potential in men, but not in women. One remaining question, then, is why are observers still sensitive to this metric when judging threat potential in female faces (see meta-analysis on perceptions of threat in Chapter 5)? One possibility is that the analysis on perceived threat potential contained too few female faces (n = 72) to accurately assess sex as a moderator. In the analysis of dominance perceptions, for example, which involved a larger set of female faces (n = 202), sex did moderate (albeit marginally) the relationship, with significant associations between perceived dominance and the male but not the female face ratio.

Another possibility is that the relationship is weaker but still significant in female faces because observers overgeneralize its use from male faces, in which it possesses some ecological validity, to female faces, in which it has less or no ecological validity (Geniole et al., 2012). Alternatively, the face ratio may be associated with less direct forms of threat behaviour (e.g., indirect aggression, ostracism, gossip) and personality traits associated with these less direct forms in women, such as coldheartedness (see results of Chapter 2). These traits and behaviours, however, are not oft investigated in studies of the face ratio and threat potential. The face ratio may also share weaker associations with threat behaviour in women than in men because of reduced variability in the female than the male face ratio. An examination of the average standard deviation for samples of female versus male face ratios in Table S5-1, however, indicates no difference (mean standard deviation in both sexes = 0.14). Future investigations in women are thus required to better address the sex specificity of the association between the face ratio and threat behaviour and determine whether or not the face ratio and sensitivity to it serves an adaptive function in female faces.

The Advertisement, Assessment, and Action Model

In Chapter 1, I proposed an Advertisement, Assessment, and Action model. Although I provided many alternative explanations for the link between facial structure and behaviour and the link between facial structure and personality in the General Discussion, a main point of this model was that the face ratio does not influence threat potential directly. Instead, it is correlated with individual differences in neural circuits and function, and neural circuits and their function underlie the personality traits and behaviours that determine threat potential. I propose that the face ratio is likely correlated with personality traits and behaviour because they all result from the same underlying biological factor, such as the effects of steroid hormones at specific phases of development. It is also possible that all of these variables, including threat potential, are influenced directly by the same underlying biological factor(s), and are thus simply correlates of one another[24]. In either case, the face ratio advertises threat potential because it is correlated with or is caused by the same factor(s) that determine threat potential. One prediction derived from this model is that if these other factors (rather than the face ratio) determine threat potential, the face ratio's association with threat potential should become weaker when some of these other variables are entered as simultaneous predictors of threat potential (as in Chapter 2).

Through the model, I also propose that observers use this metric to assess threat potential. Therefore, the face ratio should influence judgements of threat. Further, if accuracy in judgements of threat potential are possible because of the association between threat potential and the face ratio, then the individual's actual threat potential should become a weaker predictor of observers' judgements when the face ratio is entered as a simultaneous predictor in a regression model. Because the model assumes that the face ratio carries the information about threat potential, it should be a stronger predictor in the model than the individual's actual threat potential.

Finally, in the last step of this model, I propose that these assessments of threat potential modulate the actions of the observers (e.g., fight, flight, approach, avoidance). If so, then the face ratio should not only influence the observer's actions (both when this metric varies naturally and when it is manipulated experimentally), but this relationship

[24] If a dataset existed that captured all of these key variables throughout development, the assumptions of both possible models could be tested using Structural Equation Modelling.

between the face ratio and observer actions should be indirect and mediated by assessments of threat potential. In other words, the face ratio should regulate the observers' actions because men with larger face ratios are assessed as more threatening. It will be beneficial for future studies to provide further tests of these predictions, which are derived from this model. For a graphical representation of the model, see Figure 6-1. Whereas many studies have investigated links between the face ratio and actual threat behaviour (see Supplementary Tables S5-3, for examples), and between the face ratio and judgements of threat potential (see Supplementary Tables S5-7 and S5-8, for examples), fewer have investigated how this metric, and judgements based on it, regulate the actions of the observers (e.g., Efferson & Vogt, 2013; Geniole, MacDonell, & McCormick, under review; Haselhuhn et al., 2013; Hehman, Leitner, et al., 2015; MacDonell et al., 2015; Stirrat & Perrett, 2010, 2012; Valentine et al., 2014). Future studies on observer actions will thus be particularly important for better identifying the precise social function of this metric.

Conclusion

Despite its simplicity, the face ratio is an important metric for social perceptions of threat. Its broad application to a number of fields (e.g., biology, evolution, psychology, economics, anthropology, history) speaks to its value in promoting transdisciplinary research approaches, which may improve our understanding of (1) why variation in the face ratio is meaningful, (2) how observers came to be so sensitive to this metric, and (3) the social consequences of this sensitivity. Developmental approaches may also be particularly important for elucidating the mechanisms (be they biological, social, environmental) that link the face ratio to behaviour and to social judgements.

Figures

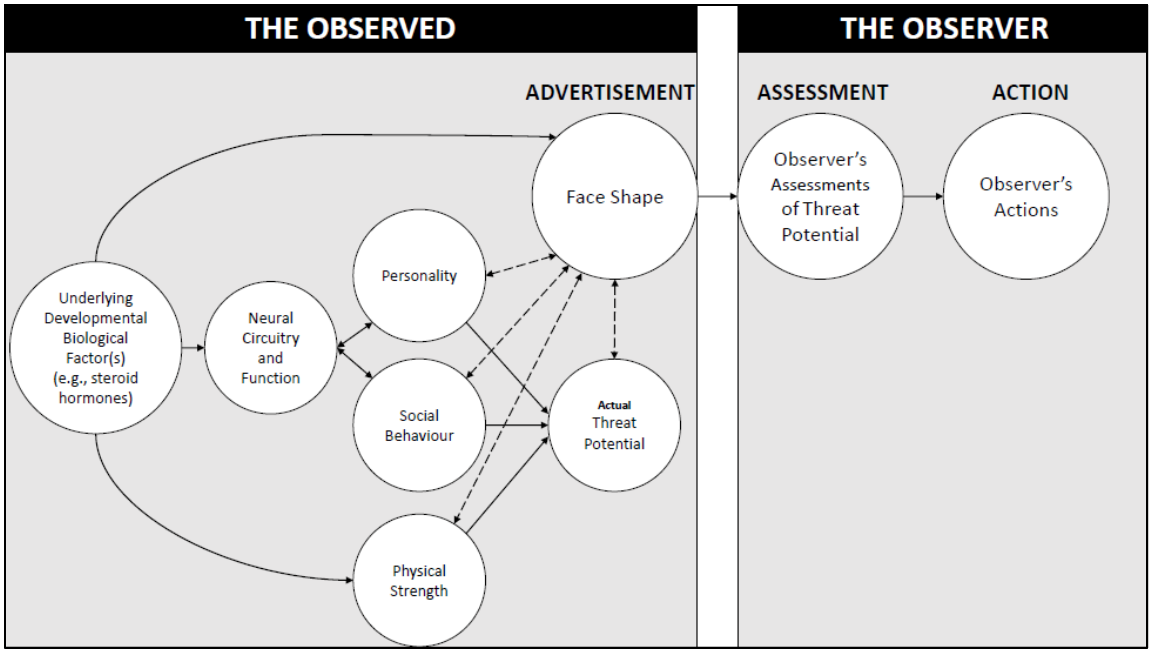

Figure 6-1. A graphical representation of the Advertisement, Assessment, and Action model. Solid lines represent causal relationships, while dashed lines represent non-causal relationships. Advertisement: The shape of the face (in this case, the face ratio) is a correlate of actual threat potential because a common underlying biological factor, such as exposure to steroid hormones at specific times in development, simultaneously influences the shape of the face, physical strength, and neural circuitry and function. Neural circuitry and function underlies personality and social behaviour, factors (in addition to physical strength) that determine actual threat potential. Because these variables all share in common this underlying biological factor, the shape of the face is also correlated with neural circuitry and function, personality, social behaviour, and physical strength, in addition to actual threat potential. Assessment: In first time (face-to-face) interactions with strangers, observers may accurately assess threat potential. This assessment is influenced by the shape of the face, and is thus accurate to the extent that

the shape of the face is a correlate of threat potential. Action: Observers respond to this assessment by modulating their actions (e.g., approach/avoid, fight/flight).

References

Alrajih, S., & Ward, J. (2014). Increased facial width-to-height ratio and perceived dominance in the faces of the UK's leading business leaders. *British Journal of Psychology*, *105*(2), 153–161.

Anderl, C., Hahn, T., Schmidt, A. K., Moldenhauer, H., Notebaert, K., Clément, C. C., & Windmann, S. (2016). Facial width-to-height ratio predicts psychopathic traits in males. *Personality and Individual Differences*, *88*, 99–101.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion*, *6*(2), 269–278.

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

Berry, D. S. (1991). Accuracy in Social Perception: Contributions of Facial and Vocal Information. *Journal of Personality and Social Psychology*, *61*(2), 298–307.

Bird, B. M., Cid Jofré, V. S., Geniole, S. N., Welker, K. M., Zilioli, S., Maestripieri, D., … Carré, J. M. (2016). Does the facial width-to-height ratio map onto variability in men's testosterone concentrations? *Evolution and Human Behavior*. doi: 10.1016/j.evolhumbehav.2016.03.004

Blakemore, S.-J. (2012). Imaging brain development: the adolescent brain. *NeuroImage*, *61*(2), 397–406.

Boduszek, D., Shevlin, M., Adamson, G., & Hyland, P. (2013). Eysenck's Personality Model and Criminal Thinking Style within a Violent and Nonviolent Offender Sample: Application of Propensity Score Analysis. *Deviant Behavior*, *34*, 483–493.

Borgi, M., & Majolo, B. (2016). Facial width-to-height ratio relates to dominance style in the genus *Macaca*. *PeerJ*, *4*, e1775.

Boshyan, J., Zebrowitz, L. A., Franklin, R. G., McCormick, C. M., & Carré, J. M. (2014). Age similarities in recognizing threat from faces and diagnostic cues. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *69*, 710-718.

Bro-Jørgensen, J., & Beeston, J. (2015). Multimodal signalling in an antelope: fluctuating facemasks and knee-clicks reveal the social status of eland bulls. *Animal Behaviour, 102*, 231–239.

Bryan, R., Perona, P., & Adolphs, R. (2012). Perspective distortion from interpersonal distance is an implicit visual cue for social judgments of faces. *PLoS ONE, 7*(9), e45301.

Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences*, *12*, 1–49.

Carré, J. M. (2014). Social status, facial structure, and assertiveness in brown capuchin monkeys. *Frontiers in Psychology, 5*, 567.

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 of the Royal Society B Biological Sciences*, *275*(1651), 2651–2656.

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.

Carrier, D. R., & Morgan, M. H. (2015). Protective buttressing of the hominin face. *Biological Reviews*, *90*(1), 330–346.

Castellani, V., Pastorelli, C., Eisenberg, N., Gerbino, M., Di Giunta, L., Ceravolo, R., & Milioni, M. (2014). Hostile, aggressive family conflict trajectories during the transition to adulthood: Associations with adolescent Big Five and emerging adulthood adjustment problems. *Journal of Adolescence*, *37*(5), 647–658.

Cherulnik, P. D., Way, J. H., Ames, S., & Hutto, D. B. (1981). Impressions of high and low Machiavellian men. *Journal of Personality*, *49*(4), 388–400.

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

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

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

Franklin, R. G., & Zebrowitz, L. A. (2013). Older adults' trait impressions of faces are sensitive to subtle resemblance to emotions. *Journal of Nonverbal Behavior*, *37,* 139-151.

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*(1), e30366.

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*(5), 1027–1043.

Goetz, S. M. 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*(11), 2329–34.

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*(1), e52317.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, *3*, 367–388.

Haselhuhn, M. P., & Wong, E. M. (2012). Bad to the bone: facial structure predicts unethical behaviour. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 571–576.

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*(8), e72259.

Haselhuhn, M. P., Wong, E. M., Ormiston, M. E., Inesi, M. E., & Galinsky, A. D. (2014). Negotiating face-to-face: Men's facial structure predicts negotiation performance. *Leadership Quarterly*, *25*(5), 835–845.

Haselton, M. G., & Buss, D. M. (2000). Error management theory: a new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology*, *78*(1), 81–91.

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., Flake, J. K., & Freeman, J. B. (2015). Static and dynamic facial cues differentially affect the consistency of social evaluations. *Personality and Social Psychology Bulletin*, *41*, 1123– 1134.

Hehman, E., Leitner, J. B., Deegan, M. P., & Gaertner, S. L. (2015). Picking teams:

When dominant facial structure is preferred. *Journal of Experimental Social Psychology*, *59*, 61–65.

Hehman, E., Leitner, J. B., & Gaertner, S. L. (2013). Enhancing static facial features increases intimidation. *Journal of Experimental Social Psychology*, *49*(4), 747–754.

Hodges-Simeon, C. R., Hanson Sobraske, K. N., Samore, T., Gurven, M., & Gaulin, S. J. C. (2016). Facial Width-To-Height Ratio (fWHR) Is Not Associated with Adolescent Testosterone Levels. *Plos One*, *11*(4), e0153083.

Holtzman, N. S. (2011). Facing a psychopath: Detecting the dark triad from emotionally-neutral faces, using prototypes from the Personality Faceaurus. *Journal of Research in Personality*, *45*(6), 648–654.

Holzleitner, I. J., & Perrett, D. I. (2016). Perception of strength from 3D faces is linked to facial cues of physique. *Evolution and Human Behavior, 37,* 217-229.

Hönekopp, J., & Watson, S. (2011). Meta-analysis of the relationship between digit-ratio 2D:4D and aggression. *Personality and Individual Differences*, *51*(4), 381–386.

Ingervall, B., & Bitsanis, E. (1987). A pilot study of the effect of masticatory muscle training on facial growth in long-face children. *The European Journal of Orthodontics*, *9*(1), 15–23.

Jenkins, R., White, D., Van Montfort, X., & Mike Burton, A. (2011). Variability in photos of the same face. *Cognition*, *121*(3), 313–323.

Jones, S. E., Miller, J. D., & Lynam, D. R. (2011). Personality, antisocial behavior, and aggression: A meta-analytic review. *Journal of Criminal Justice*, *39*(4), 329–337.

Kiliaridis, S. (1995). Masticatory muscle influence on craniofacial growth. *Acta Odontologica Scandinavica*, *53*(3), 196–202.

Kramer, R. S. S. (2015). Facial width-to-height ratio in a large sample of commonwealth games athletes. *Evolutionary Psychology*, *13*(1), 197–209.

Kramer, R. S. S. (2016). Within-person variability in men's facial width-to-height ratio. *PeerJ, 4*, e1801.

Kramer, R. S. S., Jones, A. L., & Ward, R. (2012). A lack of sexual dimorphism in width-to-height ratio in white European faces using 2D photographs, 3D scans, and anthropometry. *PloS One*, *7*(8), e42705.

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 signaling personality. *Evolution and Human Behavior*, *32*(3), 179–185.

Kramer, R. S. S., & Ward, R. (2012). Cues to personality and health in the facial appearance of chimpanzees. *Evolutionary Psychology*, *10*(2), 320–337.

Lefevre, C. E., Etchells, P. J., Howell, E. C., Clark, A. P., & Penton-voak, I. S. (2014). Facial width-to-height ratio predicts self-reported dominance and aggression in males and females , but a measure of masculinity does not. *Biology Letters*, *10,* 20140729.

Lefevre, C. E., Lewis, G. J., Bates, T. C., Dzhelyova, M., Coetzee, V., Deary, I. J., & Perrett, D. I. (2012). No evidence for sexual dimorphism of facial width-to-height ratio in four large adult samples. *Evolution and Human Behavior*, *33*(6), 623–627.

Lefevre, C. E., Lewis, G. J., Perrett, D. I., & Penke, L. (2013). Telling facial metrics: facial width is associated with testosterone levels in men. *Evolution and Human Behavior, 34*(4), 273–279.

Lefevre, C. E., Wilson, V. A. D., Morton, F. B., Brosnan, S. F., Paukner, A., & Bates, T.

C. (2014). Facial width-to-height ratio relates to alpha status and assertive personality in capuchin monkeys. *PloS One*, *9*(4), e93369.

Li, N. P., Bailey, J. M., Kenrick, D. T., & Linsenmeier, J. A. W. (2002). The necessities and luxuries of mate preferences: Testing the tradeoffs. *Journal of Personality and Social Psychology*, *82*(6), 947–955.

LoBue, V., & Rakison, D. H. (2013). What we fear most: A developmental advantage for threat-relevant stimuli. *Developmental Review*, *33*(4), 285–303.

Loehr, J., & Hara, R. B. O. (2013). Facial morphology predicts male fitness and rank but not survival in Second World War Finnish soldiers. *Biology Letters*, *9*, 20130049.

MacDonell, E. T., Geniole, S. N., & McCormick, C. M. (2015). Vulnerability versus virility: the facial cues, and sensitivity to the cues, that influence economic bargaining. Poster presented at the Canadian Society for Brain, Behaviour, and Cognitive Science. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *69*, 341.

Malatesta, C., Fiore, M., & Messina, J. (1987). Affect, personality, and facial expressive characteristics of older people. *Psychology and Aging*, *2*(1), 64–69.

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*(5), 681–690.

Marsh, A. A., Adams, R. B., & Kleck, R. E. (2005). Why do fear and anger look the way they do? Form and social function in facial expressions. *Personality & Social Psychology Bulletin*, *31*(1), 73–86.

Marsh, A. A., Cardinale, E. M., Chentsova-Dutton, Y. E., Grossman, M. R., & Krumpos, K. A. (2014). Power plays: Expressive mimicry of valid agonistic cues. *Social Psychological and Personality Science*, *5*(6), 684–690.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*(6), 255–260.

Montepare, J. M., & Dobish, H. (2003). The contribution of emotion perceptions and their overgeneralizations to trait impressions. *Journal of Nonverbal Behavior, 27*(4), 237–254.

Mueller, U., & Mazur, A. (1996). Facial dominae of west point cadets as a predictor of later military rank. *Social Forces*, *74*(3), 823–850.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(32), 11087–11092.

Paschetta, C., de Azevedo, S., Castillo, L., Martínez-Abadías, N., Hernández, M., Lieberman, D. E., & González-José, R. (2010). The influence of masticatory loading on craniofacial morphology: A test case across technological transitions in the Ohio valley. *American Journal of Physical Anthropology*, *141*, 297–314.

Puts, D. A. (2010). Beauty and the beast: mechanisms of sexual selection in humans. *Evolution and Human Behavior*, *31*(3), 157–175.

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

Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion*, *9*(2), 260–4.

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*(5), 597–604.

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 of the Royal Society B: Biological Sciences*, *276*(1656), 575–584.

Sell, A., Hone, L. S. E., & Pound, N. (2012). The importance of physical strength to human males. *Human Nature*, *23*(1), 30–44.

Serra-Negra, J. M., Paiva, S. M., Flores-Mendoza, C. E., Ramos-Jorge, M. L., & Pordeus, I. A. (2012). Association among stress, personality traits, and sleep bruxism in children. *Pediatric Dentistry*, *34*, 30E–34E.

Setchell, J. M., Smith, T., Wickings, E. J., & Knapp, L. A. (2008). Social correlates of testosterone and ornamentation in male mandrills. *Hormones and Behavior*, *54*(3), 365–72.

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*(2), 121–129.

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*(4), 298–303.

Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: male facial width and trustworthiness. *Psychological Science*, *21*(3), 349–354.

Stirrat, M., & Perrett, D. I. (2012). Face structure predicts cooperation: men with wider faces are more generous to their in-group when out-group competition is salient. *Psychological Science*, *23*(7), 718–22.

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*(5), 551–556.

Sutin, A. R., Terracciano, A., Ferrucci, L., & Costa, P. T. (2010). Teeth Grinding: Is Emotional Stability related to Bruxism? *Journal of Research in Personality*, *44*(3), 402–405.

Thomsen, L., Frankenhuis, W. E., Ingold-Smith, M., & Carey, S. (2011). Big and mighty: preverbal infants mentally represent social dominance. *Science, 331*(6016), 477–480.

Tibbetts, E. A., & Dale, J. (2004). A socially enforced signal of quality in a paper wasp. *Nature*, *432*, 218–223.

Tibbetts, E. A., & Lindsay, R. (2008). Visual signals of status and rival assessment in Polistes dominulus paper wasps. *Biology Letters*, *4*(3), 237–239.

Todorov, A. T., & Porter, J. M. (2014). Misleading first impressions: Different for different facial images of the same person. *Psychological Science*, *25*(7), 1404–1417.

Toscano, H., Schubert, T. W., & Sell, A. N. (2014). Judgments of dominance from the face track physical strength. *Evolutionary Psychology*, *12*(1), 1–18.

Trebicky, V., Havlícek, J., Roberts, S. C., Little, A. C., & Kleisner, K. (2013). Perceived aggressiveness predicts fighting performance in mixed-martial-arts fighters.

Psychological Science, *24*(9), 1664–72.

Undurraga, E. A., Eisenberg, D. T. A., Magvanjav, O., Wang, R., Leonard, W. R., McDade, T. W., … Godoy, R. A. (2010). Human's cognitive ability to assess facial cues from photographs: a study of sexual selection in the Bolivian Amazon. *PLOS ONE, 5*(6), e11027.

Valentine, K. a., Li, N. P., Penke, L., & Perrett, D. I. (2014). Judging a man by the width of his face: the role of facial ratios and dominance in mate choice at speed-dating events. *Psychological Science*, *25*(3), 806–811.

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*(2), 137–143.

Welker, K. M., Goetz, S. M. M., & Carré, J. M. (2015). Perceived and experimentally manipulated status moderates the relationship between facial structure and risk-taking. *Evolution and Human Behavior, 36*(6), 423–429.

Weston, E. M., Friday, A. E., & Liò, P. (2007). Biometric evidence that sexual selection has shaped the hominin face. *PloS One*, *2*(8), e710.

Whitehouse, A. J. O., Gilani, S. Z., Shafait, F., Mian, A., Tan, D. W., Maybery, M. T., … Eastwood, P. (2015). Prenatal testosterone exposure is related to sexually dimorphic facial morphology in adulthood. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1816), 1–9.

Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science*, *26*(8), 1325–1331.

Wilson, V., Lefevre, C. E., Morton, F. B., Brosnan, S. F., 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.

Wong, E. M., Ormiston, M. E., & Haselhuhn, M. P. (2011). A face only an investor could love: CEOs' facial structure predicts their firms' financial performance. *Psychological Science*, *22*(12), 1478–83.

Zebrowitz, L. A., Andreoletti, C., Collins, M. A., Lee, S. Y., & Blumenthal, J. (1998). Bright, bad, babyfaced boys: appearance stereotypes do not always yield self-fulfilling prophecy effects. *Journal of Personality and Social Psychology*, *75*(5), 1300–1320.

Zebrowitz, L. A., & Collins, M. A. (1997). Accurate social perception at zero acquaintance: the affordances of a Gibsonian approach. *Personality and Social Psychology Review*, *1*(3), 204–223.

Zebrowitz, L. A., Franklin, R. G. J., & Boshyan, J. (2015). Face shape and behavior: Implications of similarities in infants and adults. *Personality and Individual Differences*, *86*, 312–317.

Zebrowitz, L. A., Voinescu, L., & Collins, M. A. (1996). "Wide-eyed" and "crooked-faced": determinants of perceived and real honesty across the life span. *Personality and Social Psychology Bulletin*, *22*(12), 1258–1269.

Supplementary Materials

Chapter 3 Supplementary Analyses

Relationships between reaction times and the face ratio

In Table S3-1, I report associations among reaction times for aggression and for trustworthiness ratings, and the face ratio. When rating the faces for a second time, participants provided their ratings more quickly for larger face ratio than for smaller face ratio men. As a consequence, among participants who rated aggression first, the difference in reaction time speed between judgements of aggression and judgements of trustworthiness was smaller when rating men with larger face ratio than when rating men with smaller face ratios.

Ratio versus width and height

Using the data from Study 4 (which included judgements of 65 faces), I examined the contributions of the individual width and height components, in addition to the face ratio. The width and height components were correlated positively (r .48, $p < .001$), but shared opposing associations with judgements of aggression (width and aggression: r = .27, p = 03; height and aggression: $r = -.24$, $p = .05$). Both of these individual associations were weaker than the association between the face ratio and judgements of aggression ($r = .48$, $p < .001$).

When width and height were entered as simultaneous predictors of judgements of aggression they accounted for 25% of the variability in these judgements; both variables were significant predictors within the model (width: $\beta = .50$, $sr = .44$, $p < .001$; height: $\beta = -.48$, $sr = -.42$, $p < .001$). The ratio, when entered as the sole predictor, accounted for a similar amount of variability in judgements of aggression (23%). Adding the width and height components to a model that already contains the ratio does not significantly improve the amount of variability explained by the model (*p* = .13), nor does adding the ratio to a model that already contains the width and height components (*p* = .09). See the General Discussion for interpretation of these findings.

Chapter 3 Supplementary Tables

Table S3-1.

Bivariate correlations among the face ratio and reaction times for judgements of aggression and trustworthiness.

	FR	Rated Aggression 1st			Rated Aggression 2nd	
		AGG	TRUST	DIF	AGG	TRUST
Rated AGG 1st						
AGG	0.04					
TRUST	**-0.29**	0.05				
DIF	**-0.26**	**-0.54**	**0.82**			
Rated AGG 2nd						
AGG	**-0.26**	0.09	0.07	0.01		
TRUST	-0.09	0.03	**0.26**	0.21	-0.01	
DIF	0.13	-0.04	0.13	0.14	**-0.73**	**0.69**

Notes. FR = face ratio. AGG = aggression. TRUST = trustworthiness. DIF = difference score (trustworthiness – aggression) such that higher values indicate the face was rated faster on aggression than on trustworthiness, and lower values indicate that the faces were rated equally as fast for both judgement types, or that they were rated faster on trustworthiness than on aggression.

Chapter 4 Supplementary Materials

How do judgements of bearded versus non-bearded versions of the same faces differ within individual observers?

Paired-samples t-tests indicated that 23 participants (41%) rated the bearded versions of the faces as significantly more aggressive than the non-bearded versions whereas two (4%) rated the non-bearded versions as more aggressive than the bearded versions. Fifteen participants (27%) rated the bearded versions as more masculine than the non-bearded versions, whereas two participants (4%) rated the non-bearded versions as more masculine than the bearded versions. Eleven participants (20%) rated non-bearded versions of the faces as more attractive than the bearded versions, whereas five participants (9%) rated the bearded versions as more attractive than the non-bearded versions (all *p*s < 0.05).

How strong are the associations between the facial width-to-height ratio and each of the judgements in bearded and non-bearded versions of the same faces, within individual observers?

Judgements of aggression. Judgements of aggressiveness in the bearded and non-bearded versions of the faces were significantly correlated (critical *r*s > .396, *p*s < 0.05) for 19 (34%) individual observers (see Figure S4-1-A). Correlations between the face ratio and judgements of aggression were significant for 32 observers (57%) when the stimuli were of non-bearded versions of the faces, and for 25 observers (45%) when the stimuli were of bearded versions of the faces (see Figure S4-2-A).

Judgements of masculinity. Judgements of masculinity in the bearded and non-bearded versions were significantly correlated (critical *r*s > .396, *p*s < 0.05) for four (7%)

individual observers (see Figure S4-1-B). Judgements of masculinity were correlated with the face ratio for 19 participants (34%) for the non-bearded versions, and for no participant (0%) for the bearded versions (see Figure S4-2-B).

Judgements of attractiveness. Judgements of attractiveness in the bearded and non-bearded versions of the faces were significantly correlated (critical *r*s > .396, *p*s < 0.05) for 16 (29%) individual observers (see Figure S4-1-C). Judgements of attractiveness were significantly correlated with the face ratio for eight participants (14%) when the stimuli were of the non-bearded versions, and eight participants (14%) when the stimuli were bearded versions of the faces (see Figure S4-2-C).

Omnibus analysis on the associations between the face ratio and social judgements within individual observers. The correlations between the face ratio and each judgement for each observer were transformed using Fisher's Z and analysed in a repeated measures ANOVA with two within-subjects factors (Judgement Type: aggression, masculinity, attractiveness; Beardedness: bearded vs. non-bearded versions). The analysis revealed a main effect of Judgement Type ($F_{2,110} = 35.42$, $p < 0.001$, $\eta_p^2 = 0.39$) and of Beardedness ($F_{1,55} = 21.72$, $p < 0.001$, $\eta_p^2 = 0.28$), and an interaction of Beardedness and Judgement Type ($F_{2,110} = 8.64$, $p < 0.001$, $\eta_p^2 = 0.14$). Post-hoc analyses indicated that within the non-bearded versions of faces, the face ratio shared stronger associations with judgements of aggression than with judgements of attractiveness ($t_{55} = 6.39$, $p < 0.001$, *Cohen's d* = 0.87) and masculinity ($t_{55} = 3.84$, $p < 0.001$, *Cohen's d* = 0.52), and also stronger associations with judgements of masculinity than with judgements of attractiveness ($t_{55} = 2.96$, $p = 0.005$, *Cohen's d* = 0.41). Within the bearded versions of faces, the face ratio similarly shared stronger associations with judgements of aggression than with judgements of attractiveness ($t_{55} = 4.80$, $p < 0.001$, *Cohen's d* = 0.63) and masculinity ($t_{55} = 7.58$, $p < 0.001$, *Cohen's d* = 1.00), but unlike in the non-bearded versions of faces, the face ratio shared stronger associations with judgements of attractiveness than with judgements of masculinity ($t_{55} = 1.97$, $p = 0.05$, *Cohen's d* = 0.27).

Given the number of stimuli within bearded (n = 25) and within non-bearded stimuli sets (n = 25), a Pearson correlation coefficient value of .38 (Fisher's Z = .40) between the face ratio and a given judgement is required for statistical significance (two-tailed test). A one-sample t-test was used to determine if any of the associations with the face ratio were significantly lower than this critical value. Whereas the associations between the face ratio and judgements of aggression were not significantly lower than this critical value, neither when judgements were made of non-bearded ($t_{55} = 1.03$, p = 0.31) nor of bearded versions of the faces ($t_{55} = 1.43$, $p = 0.16$), judgements of masculinity (non-bearded: $t_{55} = 3.15$, $p = 0.003$; bearded: $t_{55} = 14.29$, $p < 0.001$) and of attractiveness (non-bearded: $t_{55} = 7.65$, $p < 0.001$; bearded: $t_{55} = 7.59$, $p < 0.001$) were significantly lower. Therefore, across individual observers, the associations between the face ratio and judgements of aggression were maintained, even when men's faces were bearded.

How does the relationship between judgements of attractiveness and facial hair differ in versions of faces intentionally selected to be low and high in the amount of facial hair?

There is inconsistency regarding the link between men's facial hair and judgements of attractiveness (reviewed in Neave & Shields, 2008). Some have proposed that the extent to which facial hair is considered attractive depends on its frequency within a distribution (Janif, Brooks, & Dixson, 2014), with bearded men becoming more attractive as the frequency of bearded men within a population decreases, and becoming less attractive as the frequency of bearded men within a population increases. Additionally, others have provided evidence that attractiveness and facial hair share a curvilinear relationship, with light stubble being preferred over full beards and shaved faces (Neave & Shields, 2008). We predicted that judgements of attractiveness would be correlated with the extent of facial hair in both set of photos but, given these aforementioned findings, that the correlation would be positive in the non-bearded versions of men's faces and negative in the bearded versions of men's faces. This hypothesis was supported: The extent of men's facial hair was positively correlated with judgements of attractiveness in non-bearded versions of the faces ($r = .53$, $p = 0.006$), but this relationship completely flipped in direction in the bearded versions of the faces ($r = -.61$, $p = 0.001$). Therefore, these findings are consistent with those of Janif and colleagues (2014), suggesting that beards are less attractive when they are rated amongst a set of men that are all bearded, but more attractive when they are rated amongst a set of men that are non-bearded. These findings also concur with the idea that attractiveness and facial hair share a curvilinear relationship. For photos of the most attractive bearded and non-bearded face, see Figure S4-3.

Source of photographs of bearded and non-bearded versions of men's faces

Table S4-1 provides links to the videos used for each of the non-bearded and bearded versions of men's faces.

Chapter 4 Supplementary Tables

Table S4-1.

Web addresses to the source of each facial photograph and the estimated time of beard growth for each stimulus face.

Stimulus ID	Non-beard Photo (video time)	Beard Photo (video time)	Estimated Duration of Beard Growth for Beard Photo[a]	Web Address
1	20s	47s	Jan 1st - Mar 28th (86 days)	http://www.youtube.com/watch?v=tWzPDNabdBs
2	46s	23s[b]	N/A	http://www.youtube.com/watch?v=xbKKsx2xC64
3	50s	49s[b]	75 days	http://www.youtube.com/watch?v=SE0L6Tef6DI
4	17s	1m 12s	>3 months (90 days)	http://www.youtube.com/watch?v=jD9UhhU1yqs
5	1s	34s	80 days	http://www.youtube.com/watch?v=6jfBnVVeKZs
6	3s	1m 15s	N/A	http://www.youtube.com/watch?v=QY8gja0_CCA
7	6s	25s	9 weeks (63 days)	http://www.youtube.com/watch?v=frziycNkDa4
8	6s	1m 4s	N/A	http://www.youtube.com/watch?v=19SPlWmkXek
9	17s	2m 44 s	N/A	http://www.youtube.com/wat

				ch?v=LuoJT9KY5nE
10	22s	37s	Day 10 – Day 28 (28 days)	http://www.youtube.com/watch?v=MEwH9pQftDI
11	3m 48s	30s[b]	≥48 days	http://www.youtube.com/watch?v=f74INcfkoa4
12	4s	36s	N/A	http://www.youtube.com/watch?v=9WyiAh0PoxM
13	2s	28s	2 weeks (14 days)	http://www.youtube.com/watch?v=FJf3grruocg
14	16s	44s	Oct 1st – Nov 25th (55 days)	http://www.youtube.com/watch?v=X_CeU56CrPI
15	4s	29s	N/A	http://www.youtube.com/watch?v=b7KrrhWS-PI
16	4s	18s	N/A	http://www.youtube.com/watch?v=jaFsl1ZJTa4
17	1s	24s	N/A	http://www.youtube.com/watch?v=zCfKJNK8hHk
18	42s	1m 46s	N/A	http://www.youtube.com/watch?v=JYwf7VFbKAY
19	49s	3m 34s	69 days	http://www.youtube.com/watch?v=0f4TO9mkms8
20	1s	14s	N/A	http://www.youtube.com/watch?v=Clci-80LPio

21	21s	19s[b]	100 days	http://www.youtube.com/watch?v=NfokEtaV2hU
22	0s	1m 19s	54 days	http://www.youtube.com/watch?v=mmSwAvDPVLg
23	1s	27s	N/A	http://www.youtube.com/watch?v=BXRaw1UbtRQ
24	0s	20s	2 months (60 days)	http://www.youtube.com/watch?v=ebsMaR-tjf0
25	12s	25s	N/A	http://www.youtube.com/watch?v=x74vq0Fh8Co
ESTIMATED BEARD GROWTH TIME RANGE			14 days – 100 days (*M* = 63.23, *SD* = 24.32)	

Note: [a]Duration of beard growth for the bearded photo was either specified by the owner of the YouTube video, or estimated based on the owner's description of the video (e.g., some owners wrote in the comments or description of the video that photos were taken, on average, xx days apart). [b]Some photo times are not ordered chronologically given that certain photos of the non-bearded versions of the faces were shown after the bearded versions of the faces. [c]this video was made to document head-hair growth; beard growth was incidentally documented.

Chapter 4 Supplementary Figures

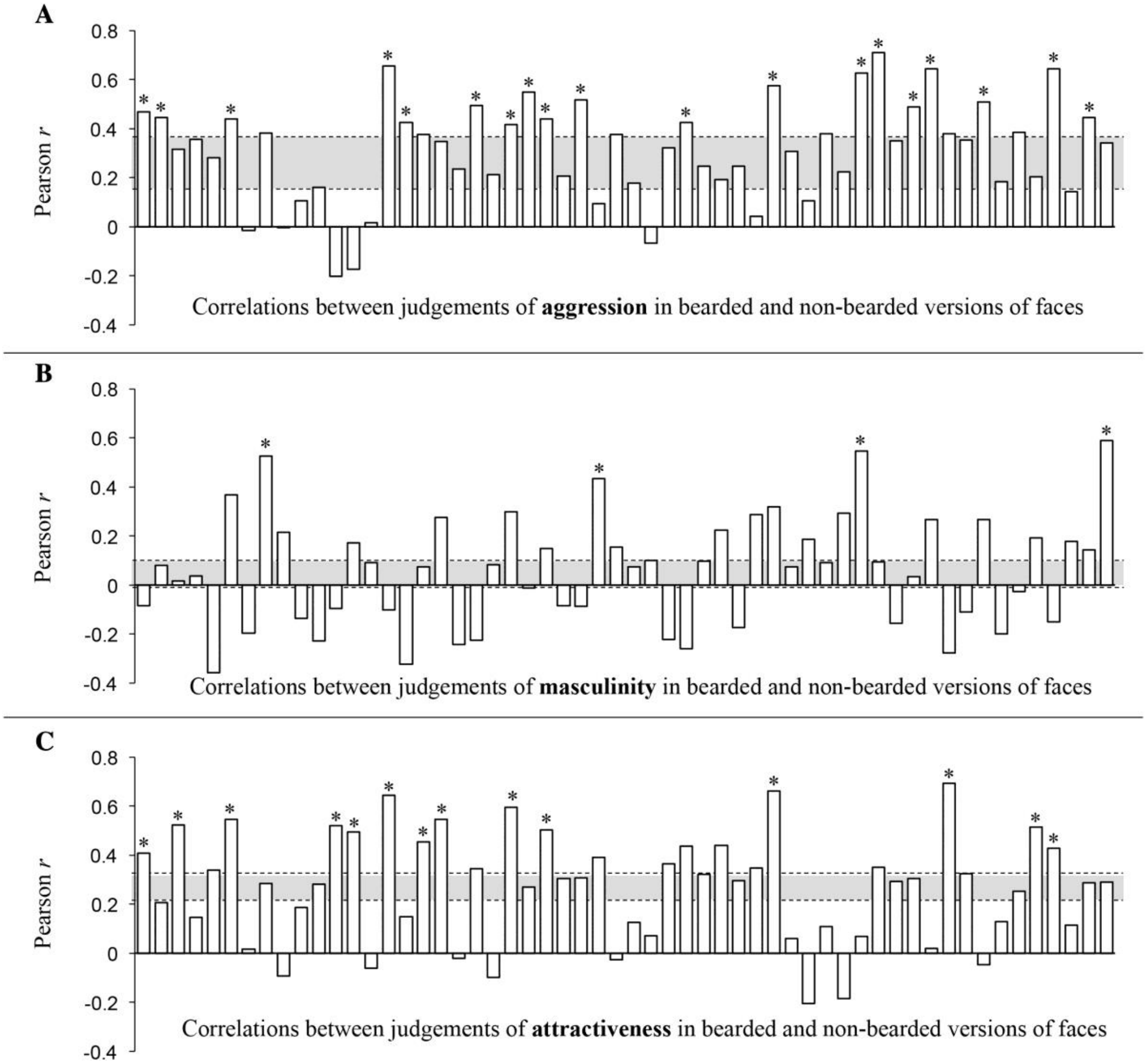

Figure S4-1. Bar graphs showing the strength of the correlation between judgements of aggression in bearded and in non-bearded versions of faces (Panel A), between judgements of masculinity in bearded and in non-bearded versions of faces (Panel B), and between judgements of attractiveness in bearded and in non-bearded versions of faces (Panel C), for each individual observer. *correlation within the individual is significant, $p < 0.05$. Shaded areas represents the 95% confidence interval obtained from a one-sampled t-test on the correlation values.

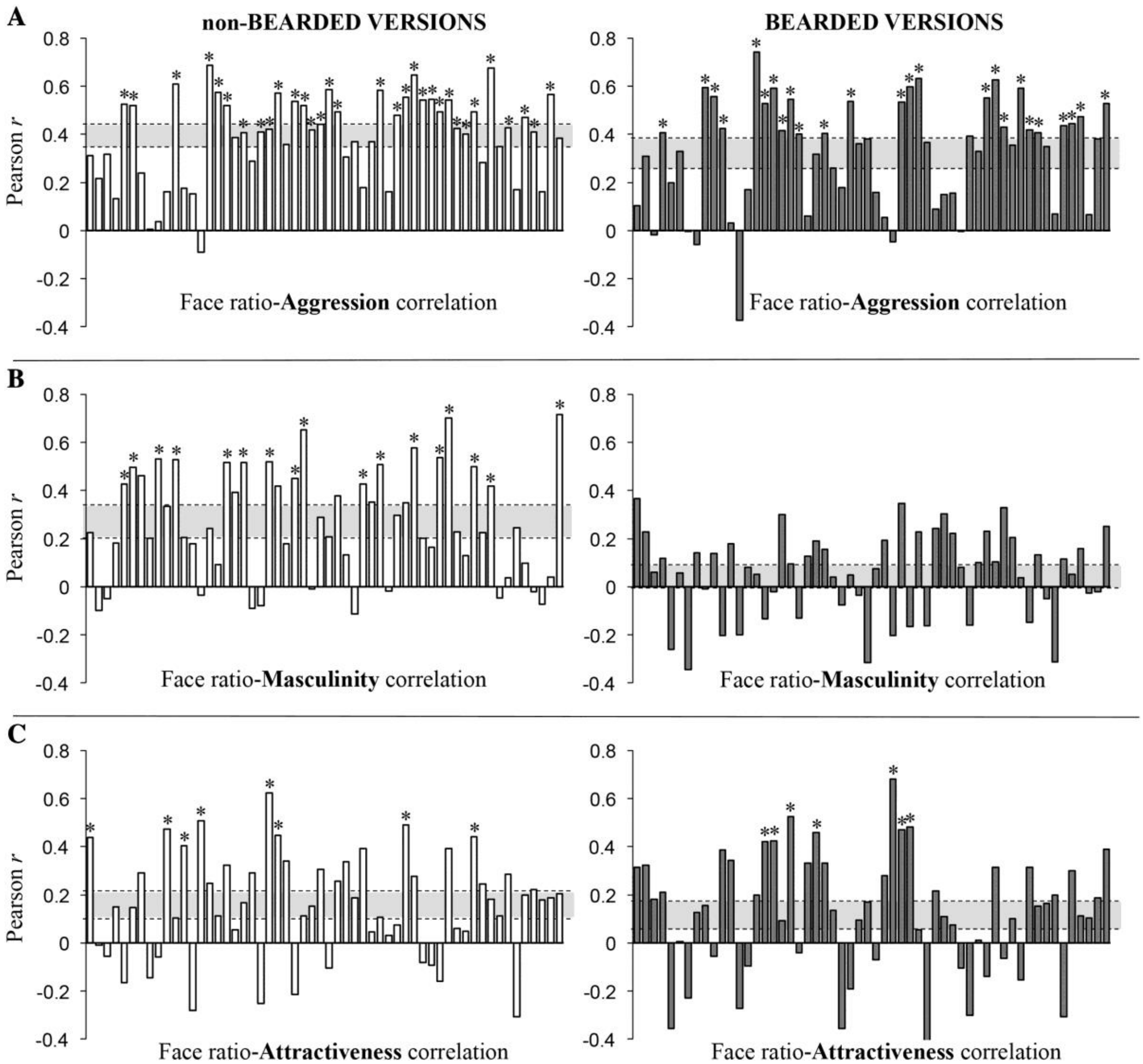

Figure S4-2. Bar graphs showing the strength of the correlation between the FWHR and judgements of aggression (Panel A), the FWHR and judgements of masculinity (Panel B), and the FWHR and judgements of attractiveness (Panel C) in non-bearded (Left Panels) and in bearded faces (Right Panels), within each individual observer ($n = 56$). *correlation within the individual is significant, $p < 0.05$. Shaded areas represent the 95% confidence interval obtained from a one-sampled t-test on the correlation values.

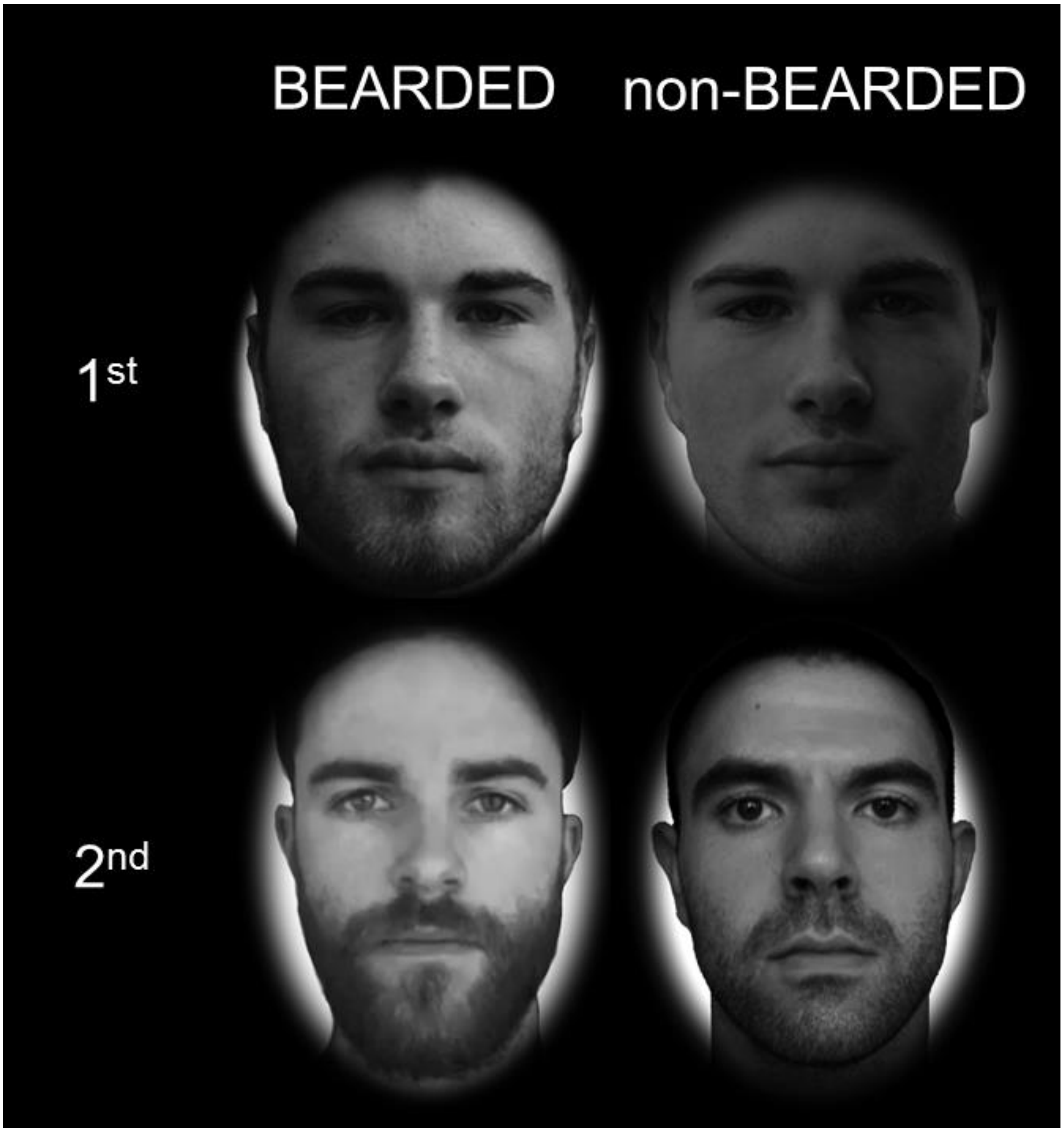

Figure S4-3. Example photos of the most and second most attractive bearded and non-bearded faces. The bearded and non-bearded version of the same face received the highest rating on attractiveness.

Chapter 4 Supplementary References

Janif, Z. J., Brooks, R. C., & Dixson, B. J. (2014). Negative frequency-dependent preferences and variation in male facial hair. *Biology letters*, *10*, 20130958. doi: 10.1098/rsbl.2013.0958

Neave, N., & Shields, K. (2008). The effects of facial hair manipulation on female perceptions of attractiveness, masculinity, and dominance in male faces. *Personality and Individual Differences*, *45*, 373–377. doi:10.1016/j.paid.2008.05.007

Chapter 5 Supplementary Materials and Methods

Additional notes regarding data extraction and effect size calculations

If more than one effect size was presented in a paper (e.g., one analysis included a covariate and another analysis did not), we used the effect size from the analysis with the fewest covariates (unless otherwise stated), thus providing the most unbiased estimate of the bivariate association between the FWHR and the variable of interest. If there were many tests of the same hypothesis using a variety of metrics that were conceptually related (explicit prejudice, implicit prejudice, motivation to respond without prejudice, Hehman, Leitner, Deegan, & Gaertner, 2013; self-reported dominance, self-reported prestige, Study 3 of Mileva, Cowan, Cobey, Knowles, & Little, 2014), we averaged the effect sizes from the study (if the *n*s used for the effects differed, we used a weighted mean). When two analyses of the same data were conducted within a data set (a correlation versus a median split and t-test), we used results from the analysis that utilized more of the data (e.g., correlations utilize more of the data array than do t-tests on a variable that has been split into low and high groups based on the median score). In studies investigating the relationship between the FWHR and social judgements, we used the effect sizes derived from analyses at the level of the stimuli rather than at the level of the individual. Specifically, some studies examine the association between the mean ratings of faces (averaged across multiple observers) and the FWHRs of the faces. Other studies examine the association between each individual observer's rating and the face ratio. If both levels of analysis were provided in a manuscript, we used the effect size from the analysis at the level of stimuli, not at the level of the observers. We also averaged the effect sizes when two highly correlated social judgements were collected from the same faces (e.g., trustworthiness and aggression, toughness and aggression).

If a given descriptive statistic (e.g., age) was reported for the sexes combined, we assumed the values to be equal in both sexes. If descriptive statistics were provided separately for men and women and we combined the sexes in an analysis, we calculated the weighted mean. If the number of men and women included in an analysis was not reported, we assumed the sexes to be equally represented (and rounded up for non-discrete numbers). When age was not reported for an effect in a college/university sample, we assumed an age of 20. If the mean age was not reported but an age range was reported, we used the midpoint of the range as the mean age of the sample. If the authors used a university/college sample, nationality was inferred based on the location of the university/college. For studies that examined sex differences in the FWHR and reported results using a variety of facial measures, we preferentially used measures obtained from 2D photos for consistency with other studies. Studies providing only 3D measures and no 2D measures (e.g., anthropometric, 3D scans, dry skulls, Gómez-Valdés et al., 2013; Stirrat, Stulp, & Pollet, 2012; Weston, Friday, & Liò, 2007) were included.

To investigate the link between FWHR and social judgements, some studies have used a continua of faces and a correlational design (examined associations between FWHRs of a variety of faces and the social judgements made for each of the faces), whereas others have compared the judgements of two or more versions of a face, with one of the versions morphed to have a larger FWHR and the other morphed to have a smaller FWHR. Results from these two types of studies were analyzed separately because of the different calculations used to determine the confidence intervals for each type of study; whereas the confidence intervals of studies with a correlational design depended on the number of faces used as stimuli, the confidence intervals of studies with an experimental design depended on the number of observers.

References for Additional Notes Regarding the Calculation of Effect Sizes

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*(1), e52317. http://doi.org/10.1371/journal.pone.0052317

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*(3), 289–96. http://doi.org/10.1177/0956797612451467

Mileva, V. R., Cowan, M. L., Cobey, K. D., Knowles, K. K., & Little, a. C. (2014). In the face of dominance: Self-perceived and other-perceived dominance are positively associated with facial-width-to-height ratio in men. *Personality and Individual Differences*, *69*, 115–118. http://doi.org/10.1016/j.paid.2014.05.019

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*(5), 551–556. http://doi.org/10.1016/j.evolhumbehav.2012.02.002

Weston, E. M., Friday, A. E., & Liò, P. (2007). Biometric evidence that sexual selection has shaped the hominin face. *PloS One*, *2*(8), e710. http://doi.org/10.1371/journal.pone.0000710

Chapter 5 Supplementary Tables of effect sizes from each meta-analysis

Table S5-1.

Effects included in the analysis of sex differences in the size of the FWHR.

Study	Sample Size		Nationality of Sample	M_{age}	Measurement Details	FWHR				Standardized Mean Difference (d), adjusted for small sample size bias
						Men		Women		
	Men	Women				*M*	*SD*	*M*	*SD*	
Carré & McCormick, 2008, Study 1	37	51	USA	18.98	2D Photo	1.860	0.130	1.800	0.100	**0.496**
Carré, Murphy, & Hariri, 2013	27	36	USA	19.39	2D Photo	1.810	0.120	1.750	0.120	0.494
Geniole, Keyes, Carré, & McCormick, 2014	146	76	North America	20.28	2D Photo	1.790	0.160	1.800	0.120	-0.068
Goetz et al., 2013, Study 1	106	113	USA	21.75	2D photo	1.718	0.122	1.703	0.148	0.110
Gómez-Valdés et al., 2013, 2D database	302	278			2D Cranial Landmark Coordinates					0.145
Gómez-Valdés et al., 2013, 3D database	401	381			3D Cranial Landmark Coordinates					0.003
Gómez-Valdés et al., 2013, Hallstat database	179	117	Austria		3D Cranial Landmark Coordinates					**0.362**
Gómez-Valdés et al., 2013, HOWELLS	2256	1156	Int		Cranial Measures					0.061

database										
Gómez-Valdés et al., 2013, Patagonia database	111	149	Argentina		Cranial Measures					0.043
Gómez-Valdés et al., 2013, Pucciarelli database	297	143			Cranial Measures					0.191
Haselhuhn & Wong, 2012, Study 1	51	45	USA	28	2D Photo	1.780	0.120	1.740	0.120	0.330
Haselhuhn & Wong, 2012, Study 2	50	53	USA	22	2D Photo	1.790	0.150	1.730	0.210	0.325
Hehman, Leitner, & Gaertner, 2013, Study 2	10	10	USA	20	2D photo	1.610	0.070	1.520	0.080	**1.140**
Huh, 2013	64	45	Republic of Korea	21.73	2D Photo	1.660	0.170	1.380	0.300	**1.197**
Huh, Yi, & Zhu, 2014	50	47	Republic of Korea	31.79	2D Photo	1.780	0.130	1.780	0.110	0.000
Denson, unpublished	42	84	Australia		2D Photo	1.861	0.1220	1.800	0.119	**0.505**
Kramer, Jones, & Ward, 2012, Study 1	138	227	German	24	2D Photo	1.850	0.110	1.870	0.110	-0.182
Kramer et al., 2012, Study 2	66	89	UK	24	2D Photo	2.010	0.160	2.030	0.140	0.133
Kramer et al., 2012, Study 3	75	105	UK	23.5	2D Photo	2.070	0.160	2.070	0.150	0.000
Lefevre, Etchells, Howell, Clark,	54	49	UK	21.59	2D Photo	2.080	0.170	2.010	0.130	**0.457**

& Penton-voak, 2014										
Lefevre et al., 2012, Sample 1	46	99	UK	20.22	2D Photo	2.120	0.182	2.170	0.157	-0.301
Lefevre et al., 2012, Sample 2	137	169	UK	83	2D Photo	2.060	0.170	2.090	0.164	-0.180
Lefevre et al., 2012, Sample 3	124	131	UK	20.34	3D Photo	1.840	0.127	1.880	0.114	**-0.331**
Lefevre et al., 2012, Sample 4	108	110	South Africa	20.22	2D Photo	2.200	0.237	2.240	0.204	-0.180
Mileva, Cowan, Cobey, Knowles, & Little, 2014, Study 1	50	50	UK	20.6	2D Photo	1.960	0.141[a]	1.960	0.141[a]	0.000
Mileva et al., 2014, Study 2	31	29	UK	21.9	2D Photo	1.980	0.111[a]	1.960	0.162[a]	0.143
Mileva et al., 2014, Study 3	21	29	UK	20.5	2D Photo	1.960	0.138[a]	1.950	0.162[a]	0.065
Özener, 2012, Study 1	230	240	Turkey	20.8	2D Photo	1.890	0.120	1.910	0.110	-0.174
Sanchez-pages, Rodriguez-ruiz, & Turiegano, 2014	147	74	Spain and UK	20.34	2D Photo	2.090	0.158[a]	2.047	0.129[a]	**0.288**
Skorska, Geniole, Vrysen, McCormick, & Bogaert, 2015	204	186	Canada	24.37	2D Photo	1.951	0.155	1.955	0.138	-0.027
Stirrat, Stulp, & Pollet, 2012	523	339	USA	36.76	Skull Measures	1.824	0.110	1.821	0.110	0.027
Weston, Friday, & Liò, 2007,	30	30	South Africa	23.69	Coordinates from Dried	1.920	0.100	1.840	0.100	**0.841**

Supporting Table S2[b]	Skulls

Effect sizes in bold were reported as significant in the corresponding manuscripts. USA = United States of America. UK = United Kingdom. Int = International sample.

[a]These values were calculated by multiplying the SEM by the square root of n.

[b]Additional data from other age groups were plotted as medians and quartiles, thus means and standard deviations were not extractable. Nevertheless, the sexual dimorphism in the FWHR is thought to emerge after puberty.

Table S5-2.

Effects included in the analysis of the relationship between the FWHR and perceptions of masculinity, restricted to studies using a correlational design and/or a continuum of faces with un-manipulated FWHRs.

Study	Observers				Stimuli					Effect size (r)[a]
	Men	Women	Nationality	M_{age}	Men	Women	Nationality	M_{age}	Display Details	
Boshyan, Zebrowitz, Franklin, McCormick, & Carré, 2013 Study 2	8	8	USA	75.6	24		USA	19.08	3000 ms, B/W	**.410**
Boshyan et al., 2013 Study 2	8	8	USA	18.8	24		USA	19.08	3000 ms, B/W	.300
Carré, McCormick, & Mondloch, 2009, Study 1	15	16	Canada	19.94	24		USA	19.08	2000 ms, B/W	**.430**
Geniole & McCormick, 2013, Study 1		29	Canada	19.41	25		Canada	19.52	1000 ms, B/W	**.420**
Geniole & McCormick, 2013, Study 1		29	Canada	19.41	54		Canada	20.32	1000 ms, Facegen, B/W	**.500**
Geniole & McCormick, 2013, Study 2		26	Canada	20.69	54		Canada	20.32	1000 ms, Facegen, B/W	**.280**
Geniole & McCormick, 2015	22	34	Canada	19.89	25				B/W, non-bearded faces[b]	**.550**
Geniole, Keyes, Mondloch, Carré, & McCormick, 2012, Study 1	10	10	Canada	20.63	24		USA	19.08	2000 ms, B/W	**.460**
Geniole et al., 2012, Study 1	10	10	Canada	20.63		31	USA	18.87	2000 ms, B/W	-.050
Geniole et al., 2012, Study 2	10	10	Canada	22.3		31	USA	18.87	2000 ms, B/W	.040

Geniole et al., 2012, Study 3	20	20	Canada	19.95	24	USA	19.08	2000 ms, B/W	.360
Sanchez-pages et al., 2014	36			31.17	147	UK, Spain	20.34		.081

Effects were from studies examining judgements of masculinity or of femininity (which we reversed). Effect sizes in bold were reported as significant in the corresponding manuscripts. USA = United States of America. UK = United Kingdom. Int = International sample.

[a]positive values indicate that individuals with larger FWHRs are judged as more masculine (or less feminine) than individuals with smaller FWHRs.

[b]the authors also examined the relationship between the FWHR and judgements of masculinity in bearded versions of the same male faces (r = .07, p = .71). We chose to only include effects from the non-bearded versions of the faces given most perceptions studies involving the FWHR have used non-bearded faces as stimuli. We note that after averaging the effect sizes, however, the mean weighted effect size is still significant (k = 12, $\bar{r}$ = .28, *95% CI* = .16 to .39, $p < .0001$; Q_{11} = 16.81, p = .11).

Table S5-3.

Effects included in the analysis on the relationship between the FWHR and threat behaviour.

Study	Sample Size		Nationality of Sample	M_{age}	Measure	Type of Measure	Type of Threat	Effect Size (r)[a]
	Men	Women						
Carré & McCormick, 2008, Study 1	37		USA	18.98	Point Subtraction Aggression Paradigm (Cherek, 1981)	Behaviour	Aggression	**.380**
Carré & McCormick, 2008, Study 1		51	USA	18.98	Point Subtraction Aggression Paradigm (Cherek, 1981)	Behaviour	Aggression	-.045[b]
Carré & McCormick, 2008, Study 2	21		Canada	20	Penalty minutes per game in varsity hockey players	Behaviour	Aggression	**.540**
Carré & McCormick, 2008, Study 3	112		Int		Penalty minutes per game in NHL players during 2007-2008 season	Behaviour	Aggression	**.300**
Carré et al., 2013	27		USA	19.39	Physical Aggression Subscale of the Buss-Perry Aggression Questionnaire (Buss & Perry, 1992)	Self-report	Aggression	-.040[b]
Carré et al., 2013		36	USA	19.39	Physical Aggression Subscale of the Buss-Perry Aggression Questionnaire (Buss & Perry, 1992)	Self-report	Aggression	.014[b]
Deaner, Goetz, Shattuck, & Schnotala, 2012	520		Int		Career penalty minutes per game of NHL players on 2011-2012 rosters	Behaviour	Aggression	.084[c]
Efferson & Vogt, 2013	41		Germany	20	Tendency to exploit another player's trust	Behaviour	Selfish and Pejorative	.369
Geniole, Keyes, et	146		North	20.28	Average of tendency to cheat (***r***	Behaviour	Selfish and	.240[‡]

al., 2014			America		**= .241**) and extent of cheating (***r* = .230**) in a lottery for a cash prize		Pejorative	
Geniole, Keyes, et al., 2014		76	North America	20.28	Average of tendency to cheat (*r* = -.042) and extent of cheating (*r* = .147) in a lottery for a cash prize	Behaviour	Selfish and Pejorative	.053[ǂ]
Goetz et al., 2013, Study 1	108		USA	21.75	Point Subtraction Aggression Paradigm (Cherek, 1981)	Behaviour	Aggression	**.187**
Goetz et al., 2013, Study 1		113	USA	21.75	Point Subtraction Aggression Paradigm (Cherek, 1981)	Behaviour	Aggression	-.008[b]
Goetz et al., 2013, Study 2	868		Int		Penalty minutes per game in NHL players during 2010-2011	Behaviour	Aggression	**.080**
Gómez-Valdés et al., 2013	163		Mexico		Difference in the size of the FWHR between the general population and weighted mean of criminal groups[d]	Behaviour	Selfish and Pejorative	-.341[e]
Haselhuhn & Wong, 2012, Study 1	51		USA	28	Explicit deception during negotiation in the Bullard House negotiation exercise (Karp, Gold, & Tan, 2006)	Behaviour	Selfish and Pejorative	**.290**
Haselhuhn & Wong, 2012, Study 1		45	USA	28	Explicit deception during negotiation in the Bullard House negotiation exercise (Karp et al., 2006)	Behaviour	Selfish and Pejorative	-.166
Haselhuhn & Wong, 2012, Study 2	50		USA	22	Inferred cheating in a lottery	Behaviour	Selfish and Pejorative	**.360**
Haselhuhn & Wong, 2012, Study 2		53	USA	22	Inferred cheating in a lottery	Behaviour	Selfish and Pejorative	-.010
Haselhuhn,	131		UK	~26[f]	Tendency for a pro-self rather	Behaviour	Selfish and	**.180**

Wong, & Ormiston, 2013, Study 1					than a prosocial resource allocation strategy		Pejorative	
Haselhuhn, Wong, Ormiston, Inesi, & Galinsky, 2014, Study 2	30 same-sex dyads		USA	20	Self-reported cooperativeness (reverse coded) in negotiation	Self-report	Selfish and Pejorative	**.410**
Hehman, Leitner, Deegan, & Gaertner, 2013, Study 1	66[g]		USA	20	Average of the relationships between the FWHR and explicit prejudice, controlling for implicit prejudice **(***n* = 70, ***sr*** **= .211)**, implicit prejudice controlling for explicit prejudice (*n* = 70, *pr* = -.038), and motivation to respond without prejudice **(reversed;** *n* = 57, ***r*** **= .295)**	Self-report[h]	Selfish and Pejorative	.147[ǂ]
Jia, van Lent, & Zeng, 2014	720		Int		CEOs risk of financial misreporting (see F-Risk values in Table 3, panel B of Jia et al., 2014)	Behaviour	Selfish and Pejorative	**.045**
Lefevre et al., 2014	54		UK	21.59	Buss-Perry Aggression Questionnaire (Buss & Perry, 1992)	Self-report	Aggression	**.270**
Lefevre et al., 2014		49	UK	21.59	Buss-Perry Aggression Questionnaire (Buss & Perry, 1992)	Self-report	Aggression	.170
Özener, 2012, Study 2	108		Turkey	20.05	Aggression Questionnaire (34 items; Buss & Warren, Buss & Warren, 2000)	Self-report	Aggression	-.001
Özener, 2012, Study 2		104	Turkey	20.74	Aggression Questionnaire (34 items; Buss & Warren, Buss	Self-report	Aggression	.051

					& Warren, 2000)			
Stirrat & Perrett, 2010	36		UK	21.6	Tendency to exploit the trust of another person for financial gain	Behaviour	Selfish and Pejorative	**.400**
Stirrat & Perrett, 2010		107	UK	21.6	Tendency to exploit the trust of another person for financial gain	Behaviour	Selfish and Pejorative	.160
Stirrat & Perrett, 2012	17		UK	20.6	Uncooperative behaviour in a public goods game	Behaviour	Selfish and Pejorative	**.500**
Třebický et al., 2015	146		Int	29.77	UFC Performance: proportion of wins to fights	Behaviour	Aggression	**.114**
Welker, Goetz, Galicia, Liphardt, & Carré, 2014	910		Int		Total fouls committed (excluding offsides) by football players in the 2010 World Cup	Behaviour	Aggression	.057
Zilioli et al., 2015, Study 1	241		Int		UFC performance: Average of FWHR correlations with total fights (***r* = .163**), number of wins (***r* = .203**), and win percentage controlling for number of fights (***r* = .139**)[i]	Behaviour	Aggression	.168[ƚ]

Effect sizes in bold were reported as significant in the corresponding manuscripts.

[ƚ]Effect size represents an average for which we did not determine statistical significance. USA = United States of America. Int = International. UK = United Kingdom. NHL = National Hockey League. [a]positive correlations indicate that individuals with larger FWHRs had greater values on the DV of interest (i.e., more threat behaviour) than those with smaller FWHRs.

[b]data obtained from one of the current manuscript's authors.

[c]Although the authors also provide penalty minutes related to fighting, we chose to use overall penalty minutes for consistency with other studies (Carré & McCormick, 2008; Goetz et al., 2013) and because it may more broadly represent the construct of aggression than penalty minutes specific to fighting. Also, the authors focused on penalty minutes to test their main hypothesis.

[d]General population (*n* = 56; mean = 1.908, *SD* = 0.125) and weighted mean and pooled *SD* of criminal groups (homicide: *n* = 58, mean = 1.838, *SD* = 0.118; robbery: *n* = 42, mean = 1.809, *SD* = 0.114; other minor fault: *n* = 7, mean = 1.765, *SD* = 0.089; total *n* of criminal groups = 107, weighted mean of criminal groups = 1.822 pooled *SD* of criminal groups = 0.115). Positive correlation value indicates a larger FWHR among criminals than the general population.

[e]effect size calculated after estimating means and standard deviations from Figure 2b in Gómez-Valdés et al. (2013).

[f]Authors did not collect information about age but drafted from a pool of students that had a mean age of 26.

[g]This n was calculated by averaging the number of participants included in each of three analyses performed by the authors (average *n* = 66).

[h]Although this measure involved an Implicit Associations Test as one of the measures of prejudice, the other two measures were self-report and thus we classified this effect as self-report rather than behavioural.

[i]we use an effect size involving a covariate here because the correlation between the FWHR and win percentage (without controlling for number of fights) was driven by fighters with relatively few fights (Zilioli et al., 2015).

Table S5-4.

Effects included in the analysis on the relationship between the FWHR and dominance.

Study	Sample Size		Nationality of Sample	M_{age}	Measure	Effect Size (r)[a]
	Men	Women				
Carré & McCormick, 2008	37		USA	18.98	Trait dominance (10 item International Personality Item Pool scale, IPIP, Goldberg et al., 2006)	-.063
Carré & McCormick, 2008		51	USA	18.98	Trait dominance (10 item IPIP, Goldberg et al., 2006)	.026
Geniole, Keyes, et al., 2014	146		North America	20.28	Fearless Dominance measured by the Psychopathic Personality Inventory-Revised (Lilienfeld & Widows, 2005)	.151[b]
Geniole, Keyes, et al., 2014		76	North America	20.28	Fearless Dominance measured by the Psychopathic Personality Inventory-Revised (Lilienfeld & Widows, 2005)	.075[b]
Haselhuhn & Wong, 2012, Study 2	50		USA	22	Psychological sense of power (Anderson & Galinsky, 2006)	**.310**
Haselhuhn & Wong, 2012, Study 2		53	USA	22	Psychological sense of power (Anderson & Galinsky, 2006)	.170
Haselhuhn et al., 2014, Study 2	30 same-sex dyads (60 men)		USA	20	Self-reported competitiveness in negotiation (How competitive do you intend to be in the upcoming negotiation, 1=not at all competitive, 7 = extremely competitive)	−.050
Lefevre et al., 2014	54		UK	21.59	Dominance (11 item IPIP, Goldberg et al., 2006)	**.290**
Lefevre et al., 2014		49	UK	21.59	Dominance (11 item IPIP, Goldberg et al., 2006)	.120
Lefevre, Lewis, Perrett, & Penke, 2013, Sample 1,	185		Germany	33.61	Competitiveness: average of the FWHRs correlations with competitiveness (r = .080) and with tendency to experience others as	.046[ǂ]

supplementary data					same-sex rivals (*r* = .011)	
Lewis, Lefevre, & Bates, 2012	29[c]		USA		President's dominance as an average of the FWHRs correlations with achievement drive (*n* =28, ***r* = .580**), forcefulness (*n* = 29, *r* = .130), and inflexibility (*n* = 29, *r* = .170)	.290[ǂ, c]
Loehr & Hara, 2013	795		Finland		Military rank at start of war (enlisted, junior officer, senior officer)	**-.152**[d]
Mileva et al., 2014, Study 2	31		UK	21.9	Trait dominance (IPIP; Goldberg, 1999) 11 items plus an additional question: "I get my own way"	**.450**
Mileva et al., 2014, Study 2		29	UK	21.9	Trait dominance (IPIP; Goldberg, 1999) 11 items plus an additional question: "I get my own way"	.160
Mileva et al., 2014, Study 3	21		UK	20.5	Self-reported dominance [average of dominance (***r* = .510**) and prestige (*r* = -.020)] (Prestige-Dominance Questionnaire, Cheng, Tracy, & Henrich, 2010)	.245[ǂ]
Mileva et al., 2014, Study 3		29	UK	20.5	Self-reported dominance [average of dominance (*r* = .030) and prestige (*r* = -.020)] (Prestige-Dominance Questionnaire, Cheng et al., 2010)	.005[ǂ]
Valentine, Li, Penke, & Perrett, 2014	78		Germany	26.5	Assured-dominant and unassured-submissive (reverse-scored) (German Revised Interpersonal Adjective Subcales, Ostendorf, 2001) controlling for adiposity	-.050

Effect sizes in bold were reported as significant in the corresponding manuscript. When analyses were conducted at the level of a dyad rather than at the level of the individual, we counted each dyad as one participant. USA = United States of America. UK = United Kingdom. IPIP = International Personality Item Pool.

[ǂ]Effect size represents an average for which we did not determine statistical significance.

[a]positive values indicate that individuals with larger FWHRs were more dominant on the corresponding measure than were individuals with smaller FWHRs.

[b]these bivariate effects were not reported in the paper but are provided here by authors.

[c]These values reflect the average n and the effect size weighted by the number of participants.

[d]we accessed the online database to obtain this bivariate correlation.

Table S5-5.

Effects included in the analysis of the relationship between the FWHR and success in business-related outcomes.

Study	Sample Size Men	Nationality of Sample	M_{age}	Measure	Effect Size (r)[a]
Alrajih & Ward, 2014	186	UK	52.5	CEO (n = 93) vs controls matched on ethnicity and (approximate) age (n = 93)	**.481**
Yang, Chao, Fabiansson, & Denson, unpublished manuscript	86 groups (244 individuals)	China	20	Value claimed in negotiation	.060
Haselhuhn et al., 2014, Study 1	23 same-sex dyads (46 men)	USA	20	Seller's negotiation performance (sale price) in 23 same sex dyads (n = 46)	**.430**[b]
Haselhuhn et al., 2014, Study 2	30 same-sex dyads (60 men)	USA	20	Negotiation for a signing bonus	**.420**[c]
Haselhuhn et al., 2014, Study 3	34 same- sex dyads (70 men)[d]	USA	20	Ability to legitimately (within the rules of the bargaining exercise) come to an agreement between buyer (who can't afford price) and seller (who can't make price any lower).	**-.355**[e]
Wong, Ormiston, & Haselhuhn, 2011	55	USA	55.7	CEO's return on assets for corresponding firm (average of 2003 and 2004)	.230

Effect sizes in bold were reported as significant in the corresponding manuscripts. When analyses were conducted at the level of a dyad rather than at the level of the individual, we counted each dyad as one participant. USA = United States of America. UK = United Kingdom. This analysis does not include effects from Jia et al. (2014) given difficulty and inconsistency among the coders in determining which effects should be included in the analysis and difficulty in determining the construct each measure represented. A study by Mayew, Parsons, and Venkatachalam (2013) was not included because the authors did not provide numerical values. [a]Positive numbers indicate that individuals with larger FWHRs perform better than individuals with smaller FWHRs. [b]This effect reflects the total FWHR size (seller FWHR/ buyer FWHR + seller FWHR). [c](candidate FWHR / candidate + recruiter FWHR) higher number indicates candidate has higher FWHR than recruiter. [d]two men were assigned to play the same role (both played as one member of one of the dyadic interactions. [e]This effect size was obtained by converting the χ^2 value to an *r* given that other conversion techniques returned unrealistically high *r* values (*r*s > .90).

Table S5-6.

Effects included in the analysis of the relationship between the FWHR and sports performance.

Study	Sample Size Men	Nationality of Sample	M_{age}	Measure	Effect Size (r)[a]
Třebický et al., 2015	146	Int	29.77	UFC Performance: proportion of wins to fights	**.114**
Tsujimura & Banissy, 2013, Study 1	104	Japan	28.91	Baseball batting performance in Japanese Central League Pennant baseball (2011). Average of FWHR correlations with batting average (*r* = .171), number of home runs (***r* = .250**), slugging percentage (***r* = .206**), hits (*r* = .112), runs-batted-in (*r* = .176), and on base percentage (*r* = .137).	.175ⱡ
Welker et al., 2014, defenders, midfielders, forwards	910	Int		Average of correlation between FWHR and total goals (*r* = .026) and of FWHR and total assists (*r* = .002) controlling for total games played, height, weight, fouls committed against the player, and player position (defender, midfielder, forward)	.014ⱡ
Welker et al., 2014, forwards only	211	Int		Average of correlation between FWHR and total goals (***r* = .152**) and of FWHR and total assists (***r* = .136**) among forwards, controlling for total games played, height, weight, fouls committed against the player	.144ⱡ
Zilioli et al., 2015, Study 1	241	Int		UFC fight performance: Average of FWHR correlation with total fights (***r* = .163**), number of wins (***r* = .203**), and win percentage controlling for number of fights (***r* = .139**)[b]	.168ⱡ

Correlations in bold were reported as significant in the corresponding manuscripts.

[ǂ]Effect size represents an average for which we did not determine statistical significance. Int = International sample. UFC = Ultimate Fighting Championships. [a]positive values indicate that individuals with larger FWHRs have better sports performance than those with lower FWHRs.

[b]we use an effect size involving a covariate here because the correlation between the FWHR and win percentage (without controlling for number of fights) was driven by fighters with relatively few fights.

Table S5-7.

Effects included in the analysis of the relationship between the FWHR and perceptions of threat, restricted to studies using a correlational design and/or a continuum of faces with un-manipulated FWHRs.

Study	Observers				Stimuli					Judgement[a]	Effect Size (r)[b]
	Men	Women	Nation-ality of sample	M_{age}	Men	Women	Nation-ality of sample	M_{age}	Display Details		
Alrajih & Ward, 2014	4[c]	6[c]	UK	21.5	153[d]		UK	52.5	Colour, 79 CEOs (m actual age = 52.5) and 74 controls (matched ethnicity, age, facial hair, glasses)	A: *r* = .060; T: *r* = .080	.070[‡]
Boshyan et al., 2013, Study 2	20[e]	20[e]	USA	18.8[f]	24		USA	19.08	3000 ms, B/W	A	**.730**
Boshyan et al., 2013, Study 2	18[g]	18[g]	USA	75.6[f]	24		USA	19.08	3000 ms, B/W	A	**.470**
Carré et al., 2009, Study 1	15	16	Canada	19.94	24		USA	19.08	2000 ms, B/W	A: ***r* = .590**; T: *r* = .450	.520[‡]
Carré et al., 2009, Study		16	Canada	19.38	24		USA	19.08	39 ms, B/W; 39	A, 39 ms display: ***r* =**	.685[‡,h]

2 and Carré, Morrissey, Mondloch, & McCormick, 2010, Exp 1B								ms, B/W, Blurred	**.700**; A, 39 ms and blurred display: ***r* = .670**	
Carré et al., 2010, Exp 1A	8	8	Canada	23.81	24	USA	19.08	2000 ms, B/W	A, chin/forehead crop display: ***r* = .790**; A, side crop display: ***r* = .810**; A, Blurred display: ***r* = .670**	.757[ɫ,h]
Carré et al., 2010, Exp 1C	1	9	Canada	23.85	24	USA	19.08	2000 ms, B/W, Blurred	A	**.450**
Efferson & Vogt, 2013	13	15	Ger	20	54	Ger			T	.132
Geniole & McCormick, 2013, Study 2[i]		26	Canada	20.69	54	Canada	20.32	1000 ms, Facegen, B/W	A	**.640**
Geniole & McCormick, 2015	22	34	Canada	19.89	25			B/W, non-bearded faces[j]	A	**.660**
Geniole et al., 2012,	10	10	Canada	20.63	24	USA	19.08	2000 ms, B/W	A	**.710**

Study 1											
Geniole et al., 2012, Study 1	10	10	Canada	20.63		31	USA	18.87	2000 ms, B/W	A	**.400**
Geniole et al., 2012, Study 2	10	10	Canada	22.3		31	USA	18.87	2000 ms, B/W	A	**.440**
Geniole, Molnar, Carré, & McCormick, 2014, footnote #3	4	36	Canada	20.23	65		North America	19.30	1000 ms display, B/W	A	**.420**
Geniole, Molnar, et al., 2014, Study 1	5	28	Canada	20.6	54		Canada	20.32	Facegen, 1000 ms, B/W	A: ***r* = .620**; T: ***r* = .470**	.545[ǂ]
Geniole, Molnar, et al., 2014, Study 1	5	29	Canada	20.6	25		Canada	19.52	1000 ms, B/W	A: ***r* = .610**; T: ***r* = .520**	.565[ǂ]
Geniole, Molnar, et al., 2014, Study 2[k]	12	12	Canada	19.58	22		Canada	19.52	Facial hair, 1000 ms, B/W	A: ***r* = .450**; T: *r* = .280	.365[ǂ]
Geniole, Molnar, et al., 2014, Study 4	8	32	Canada	19.38	65		North America	19.30	Photos displayed until response was made, B/W	A: ***r* = .480**; T: ***r* = .400**	.440[ǂ]

Hehman, Leitner, & Freeman, 2014, Study 3	8	8			60		USA	45	10 white faces from each decade (20s-70s)	F	**.809**
Hehman, Leitner, & Gaertner, 2013, Study 2	51	50	USA	20	10	10	USA		Head at baseline vs tilted up vs tilted down	A	**.572**
Hehman, Leitner, Deegan, et al., 2013, Pilot data, footnote 2	25	25	USA	20	20[l]		USA[l]			P	**.534**
Hehman, Leitner, Deegan, et al., 2013, Study 2	28[m]	74[m]	USA	20	20		USA		5 faces from each quartile of FWHR distribution from Study 1	P	**.689**
Hehman, Leitner, Deegan, et al., 2013, Study 3	21[n]	26[n]	USA	20	20		USA			P	**.747**
Kleisner, Priplatova, Frost, &	43[o]	62[o]	Czech Republic	23.1	40		Czech Republic	20.8	Colour	T	**.346**

Flegr, 2013										
Sanchez-pages et al., 2014	11[p]	11[p]		27.48	147	UK, Spain	20.34		T	**.339**
Short et al., 2012 Exp1	8	8	Canada	21.5	24	USA	19.08	2000 ms, B/W	A	**.640**
Short et al., 2012 Exp1	8	8	Canada	21.5	24	China		2000 ms, B/W	A	**.610**
Short et al., 2012 Exp1	8	8	China	21.5	24	USA	19.08	2000 ms, B/W	A	**.560**
Short et al., 2012 Exp1	8	8	China	21.5	24	China		2000 ms, B/W	A	**.470**
Short et al., 2012 Exp2	8	8	Canada	8	24	USA	19.08	2000 ms, B/W	A	**.450**
Short et al., 2012 Exp2	8	8	Canada	8	24	China		2000 ms, B/W	A	**.500**
Short et al., 2012 Exp2	8	8	China	8	24	USA	19.08	2000 ms, B/W	A	**.460**
Short et al., 2012 Exp2	8	8	China	8	24	China		2000 ms, B/W	A	.340
Stirrat & Perrett, 2010, Study 2	17	45	UK	20.32	67	UK	20.8		T	**.396**
Třebický et al., 2015	216	402	Czech Republic	26.46	146	Int	29.77		A	**.161**[q]
Třebický et al., 2015	98	180	Czech Republic	27.53	146	Int	29.77		F	**.157**[q]
Valentine et al., 2014[r]	16	15	Ger	26.39	78	Ger	26.5		A	.150

Zilioli et al., 2015, Study 2b	16	16	Canada	21.25	48		Int	Rated individuals	F	**.460**

Effect sizes in bold were reported as significant in the corresponding manuscripts. [†]Effect size represents an average for which we did not determine statistical significance. We did not include studies if the photos involved faces intentionally posed in non-neutral expressions (e.g., we did not include effects from Marsh, Cardinale, Chentsova-Dutton, Grossman, and Krumpos, 2014). USA = United States of America. UK = United Kingdom. Ger = Germany. Int = International sample. A = judgements of aggression. F = judgements of formidability (e.g., How tough does this person look?). T = judgements of trustworthiness, reversed. P = judgements of prejudice. B/W = Black and white. [a]If multiple effects are provided in a given cell, they were averaged to form a single effect size.

[b]positive values indicate that individuals with larger FWHRs are judged as more threatening than individuals with lower FWHRs.

[c]Sample size of participants based on 8 men, 12 women, who only made three rating types each. Given 6 ratings total, this sample size was estimated to be 4 men and 6 women for each rating.

[d]Although the stimuli was 93 control and 93 CEO faces, thirty-three (14 CEOs) faces were removed because they were recognized by participants.

[e]24 young adults from accuracy condition in Study 1 + 16 new young adults; 8 men, 8 women; sexes assumed to be equally represented.

[f]authors combined multiple samples for this analysis, thus making it difficult to determine the mean age. Here, we report the mean age of the new observers added to the analysis.

[g]35 total (19 old adults from the accuracy condition in Study 1 in addition to 16 new old adults); sexes assumed to be equally represented.

[h]We did not include the correlation obtained from a condition in which the faces were scrambled given this presentation strategy was meant to disrupt the relationship.

[i]Study 1 was not included given it involved a subset of the participants used in Study 1 of Geniole and colleagues Geniole, Molnar, et al., 2014.

[j]the authors also examined the relationship between the FWHR and judgements of aggression in bearded versions of the same male faces (r = .59, p = .002). We chose to only include effects from the non-bearded versions of the faces given most perception studies involving the FWHR have used non-bearded faces as stimuli. We note that averaging the effect sizes, however, leads to a similar mean weighted effect size in the final analysis (k = 38, $\bar{r}$ = .48, *95% CI* = .41 to .55, p < .0001; Q_{37} = 124.55, p < .0001).

[k]Study 3 of Geniole, Molnar, et al. (2014) was a reanalysis of data from Carré et al. (2009) and Geniole et al. (2012) and was thus not included.

[l]Not reported but assumed the same as Study 2.

[m]one participant of an unspecified sex was removed because he or she recognized one of the stimulus faces. Here we report the *n* before the removal given this ambiguity regarding sex.

[n]two participants of an unspecified sex were removed because they recognized one of the stimulus faces. Here we report the *n* before the removal of the participants given this ambiguity regarding sex.

[o]Number of each sex inferred based on distribution of total sample (142 women, 98 men) from which this subset of 105 was derived.

[p]Although the authors also reported results after splitting the faces into wide and narrow FWHR groups and performing a t-test on the mean differences in judgements between the two groups, we used the effect size based on the correlation between the FWHR and judgements across all faces given this correlation analysis utilizes every data point.

[q]aggressiveness and fighting ability were not averaged in this case because a separate set of raters completed each judgement.

[r]The authors also have the rating "interest in person as a friend", which we chose not to include here because a lack of interest in friendship does not necessarily imply threat, especially given that the friendship ratings were made in the context of speed dating appraisals.

Table S5-8.

Effects included in the analysis of the relationship between the FWHR and perceptions of threat, restricted to studies using an experimental design wherein the FWHR was manipulated to appear larger or smaller.

Study	Observers				Stimuli					Judgement[b]	St. Mean Dif (*d*), adj[c]
	M	W	Nationality	M_{age}	M[a]	W[a]	Nationality	M_{age}	Transformation Details		
Bashir & Rule, 2014	18	47	Canada	21.34	1		Canada		1 face, Morphed to have high or low FWHR	T	**-0.566**
Hehman et al., 2014, Study 1	22	22			30			44	30 faces randomly generated and modified in Facegen to look young (~18) middle-aged (~40), and old (~70), each age morphed to a high and low FWHR	A	**3.804**
Lefevre & Lewis, 2013, Study 1	34	68	UK	25.91	12		UK		12 composites x 2 prototype sets, transformed 25%, 37.5%, and 50% in shape difference of high and low FWHRs	A	**0.208**

Lefevre & Lewis, 2013, Study 2	68	190	UK	24.10	12		UK		12 composites x 1 prototype sets, transformed 37.5% and 50% in shape difference of high and low FWHRs	A, 37.5% transform: **unadjusted *d* = .305**; A, 50% transform: **unadjusted *d* = .468**	0.386[ǂ]
Lefevre & Lewis, 2013, Study 2	68	190	UK	24.1		15	UK		15 composites x 1 prototype sets, transformed 37.5% and 50% in shape difference of high and low FWHRs	A, 37.5% transform: **unadjusted *d* = 0.524**; A, 50% transform: **unadjusted *d* = 0.643**	0.582[ǂ]
Stirrat & Perrett, 2010, Study 3	77	208		23.20	12		UK	22.50	12 faces x 2 warp directions on FWHR x 3 transformation sets. 3 groups of participants rated each transformation set	T	**0.319**
Wang, Geigel, & Herbert, 2013	37	48		38	2				2 avatars, each warped to have small, medium, and large FWHRs	A	**-0.783**
Wang et al., 2013	37	48		38		2			2 avatars, each warped to have small, medium, and	A	**-0.516**

								large FWHRs		
Zilioli et al., 2015, Study 2a	20	16	Canada	21.74	15	Int		15 wide, 15 narrow composites	F[d]	**0.557**
Zilioli et al., 2015, Study 2a all white faces[e]	20	20	Canada	20.15	12	Int		12 wide, 12 narrow composites, all white faces	F	**0.794**
Zilioli et al., 2015, Study 3	66	58	Canada	21.30	12	Int	22.50	Photos from Stirrat & Perrett, 2010	F	**0.328**

Effect sizes in bold were reported as significant in the corresponding manuscripts.

St. Mean Dif (*d*), adj = Standardized mean difference, adjusted for small sample size bias.

[†]Effect size represents an average for which we did not determine statistical significance. USA = United States of America. UK = United Kingdom. Int = International sample.

[a]Given ambiguity regarding the exact number of face pairs included in each analysis, especially after we averaged across some effects within a given study, we only report the number of base images or base composites used to create the images.

[b]If multiple effects are provided in a given cell, they were averaged to form a single effect size.

[c]positive values indicate that faces manipulated to have larger FWHRs were judged as more threatening than those manipulated to have smaller FWHRs.

[d]After each face was rated individually, the authors also showed the pairs of faces side by side and had participants pick the one they believed to be tougher in a physical fight. Here, we use the D from the ratings in which each face of the pair was rated individually and then each image was compared to its counter-face, given such individual ratings provide more data unique to each facial identity.

[e]The authors also conducted a study wherein they had the individual faces, which were used to make the high and low FWHR composites, rated individually. The authors reported the correlation between the FWHR and the judgements and also reported the t-test value when the faces with high FWHRs were compared to those with smaller FWHRs. Given the second analysis utilizes less data than the first, we only used the first, which was included in the meta-analysis examining the link between the FWHR and perceptions of threat among studies that used a correlational design.

Table S5-9.

Effects included in the analysis of the relationship between the FWHR and perceptions of dominance, restricted to studies using a correlational design and/or a continuum of faces with un-manipulated FWHRs.

Study	Observers				Stimuli					Judgement	Effect Size (r)[a]
	M	W	National-ity	M_{age}	M	W	National-ity	M_{age}	Display Details		
Alrajih & Ward, 2014	4[b]	6[b]	UK	21.5	153[c]		UK	~52.5	Colour, 79 CEOs 74 controls (matched on ethnicity, age, facial hair, glasses)	D	**.280**
Burton & Rule, 2013, Study 3, (average of ratings from Study 2A and S2B)[g]	34[d]	46[d]	USA	35[e]	50[f]	69[f]	Canada	23	B/W	D	**.290**
Carré et al., 2009, Study 1	15	16	Canada	19.94	24		USA	19.08	2000 ms, B/W	D	**.540**
Hehman et al., 2014, Study 3	12	12			60		USA	45	10 white faces from each decade (20s-70s)	SP	**.720**
Little, Mileva et al., 2014, Study 1	10	9		26.4	50		UK	20.6		D	**.340**
Mileva et al., 2014, Study 1	10	9		26.4		50	UK	20.6		D	-.110

Re et al., 2013	11	11	UK	25.32	47	83		23.84	Faces retrieved from www.3d.sk.	L	-.040
Valentine et al., 2014	11[h]	44[h]	Germany	19.6	77		Germany	26.5		D	**.220**

Effect sizes in bold were reported as significant in the corresponding manuscripts. D = dominance. SP = social power. L = leadership (e.g., "How good of a leader do you think this person is?"). USA = United States of America. UK = United Kingdom. Int = International sample.

[a]positive values indicate that individuals with larger FWHRs are judged as more dominant than individuals with lower FWHRs.

[b]Sample size was estimated as these values given 8 men and 12 women were reported to have made three of six rating types each.

[c]Although the stimuli was 93 control and 93 CEO faces, thirty-three (14 CEOs) faces were removed because they were recognized by participants.

[d]These values were derived based on the estimate that the sample was 42 percent male.

[e]This value was the mean age reported for Study 2A; there was no age reported for 2B.

[f]two target faces were excluded; the sex of the targets was unspecified so we report the number of faces before exclusion.

[g]We do not provide separate effects from studies 2A and 2B because the relationships with the FWHR were not reported in these individual studies; the relationship with the FWHR was only reported after the ratings from these studies were averaged.

[h]The sample included one observer of an unknown sex. We only report the *n* for the number of participants of a known sex.

Table S5-10.

Effects included in the analysis of the relationship between the FWHR and perceptions of attractiveness, restricted to studies using a correlational design and/or a continuum of faces with un-manipulated FWHRs.

Study	Observers				Stimuli					Judgement[a]	Effect Size (*r*)[b]
	Men	Women	Nationality	M_{age}	Men	Women	Nationality	M_{age}	Display Details		
Alrajih & Ward, 2014	4[c]	6[c]	UK	21.5	153[d]		UK	~52.5	Colour, 79 CEOs (m actual age = 52.5) and 74 controls (matched on ethnicity, age, facial hair, glasses)	ATT	-.100
Boshyan et al., 2013 Study 2	8	8	USA	75.6	24		USA	19.08	3000 ms, B/W	ATT	-.320
Boshyan et al., 2013 Study 2	8	8	USA	18.8	24		USA	19.08	3000 ms, B/W	ATT	**-.470**
Carré et al., 2009, Study 1	15	16	Canada	19.94	24		USA	19.08	2000 ms, B/W	ATT	-.224
Geniole & McCormick, 2013, Study 1		30	Canada	19.43	25		Canada	19.52	1000 ms, B/W	ATT: ***r* = -.390**); ST: ***r* = -.440**; LT: ***r* = -**	-.433[ǂ]

										.470	
Geniole & McCormick, 2013, Study 1		30	Canada	19.43	54		Canada	20.32	1000 ms, Facegen, B/W	ATT: ***r* = -.700**; ST: ***r* = -.690**; LT: ***r* = -.770**	-.720[ǂ]
Geniole & McCormick, 2013, Study 2		26	Canada	20.69	54		Canada	20.32	1000 ms, Facegen, B/W	ATT	**-.550**
Geniole & McCormick, 2015	22	34	Canada	19.89	25				B/W, non-bearded faces[e]	ATT	.330
Geniole et al., 2012, Study 1	10	10	Canada	20.63	24		USA	19.08	2000 ms, B/W	ATT	-.240
Geniole et al., 2012, Study 1	10	10	Canada	20.63		31	USA	18.87	2000 ms, B/W	ATT	-.010
Geniole et al., 2012, Study 2	10	10	Canada	22.3		31	USA	18.87	2000 ms, B/W	ATT	.060
Haselhuhn et al., 2014, Study 4	2	2	USA		107		USA			ATT	-.080
Stirrat & Perrett, 2010, Study 2	17	24	UK	21	67		UK	20.8		ATT	**-.320**
Valentine et al., 2014		15	Germany	22.67	78		Germany	26.5		ATT[f]	-.180

Effect sizes in bold were reported as significant in the corresponding manuscripts.

[ŧ]Effect size represents an average for which we did not determine statistical significance. ATT = Attractiveness or unattractiveness (reversed). ST = Short-term desirability. LT = Long-term desirability. USA = United States of America. UK = United Kingdom. Int = International sample.

[a]If multiple effects are provided in a given cell, they were averaged to form a single effect size.

[b]positive values indicate that individuals with larger FWHRs are judged as more attractive than individuals with smaller FWHRs.

[c]Sample size of participants based on 8 men, 12 women, who only made three rating types each. Given 6 ratings total, this sample size was estimated to be 4 men and 6 women for each rating.

[d]Although the stimuli was 93 control and 93 CEO faces, thirty-three (14 CEOs) faces were removed because they were recognized by participants.

[e]In this manuscript the authors also examined the relationship between the FWHR and judgements of attractiveness in bearded versions of the same male faces ($r = .25$, $p = .23$). We chose to only include effects from the non-bearded versions of the faces given most perceptions studies involving the FWHR have used non-bearded faces as stimuli. We note that averaging the effect sizes, however, leads to the same mean weighted effect size in the final analysis ($k = 14$, $\bar{r} = -.26$, *95% CI* = -.40 to -.11, $p = .001$; $Q_{13} = 49.59$, $p < .0001$).

[f]Some participants also speed-dated the men and, afterwards, provided ratings of interest in the men for short- and long-term relationships. Data from these social judgements were not included given that these effects likely reflect impressions based on social interactions rather than the facial structure of the men.

Table S5-11.

Effects included in the analysis on the relationship between the FWHR and BMI.

Study	Sample Size Men	Sample Size Women	Nationality of Sample	M_{age}	Effect Size (r)[a]
Coetzee, Chen, Perrett, & Stephen, 2010[b], Study 1, Caucasian Set A		42	UK	20.9	**.480**
Coetzee et al., 2010[b], Study 1, Caucasian Set B		52	UK	19.9	**.390**
Coetzee et al., 2010[b], Study 1, African Set A		51	South Africa	19.8	.270
Coetzee et al., 2010[b], Study 1, African Set B		48	South Africa	19.6	**.330**
Coetzee et al., 2010[b], Study 1, Caucasian Set A	41		UK	21.3	**.330**
Coetzee et al., 2010[b], Study 1, Caucasian Set B	54		UK	20.4	.120
Coetzee et al., 2010[b], Study 1, African Set A	45		South Africa	21.2	.150
Coetzee et al., 2010[b], Study 1, African Set B	47		South Africa	19.9	.100
Lefevre et al., 2012, Sample 1	46	99	UK	20.22	**.270**
Lefevre et al., 2012, Sample 2	137	169	UK	83	**.230**
Lefevre et al., 2012, Sample 3	91	98	UK	20.34	**.400**
Lefevre et al., 2012, Sample 4	108	110	South Africa	20.22	**.230**

Kramer et al., 2012, Study 3		105	UK	23.5	**.430**
Kramer et al., 2012, Study 3	75		UK	23.5	**.520**
Lefevre et al., 2013, Sample 1, from supplementary data	188		Germany	33.6	**.345**
Lefevre et al., 2013, Sample 2, from supplementary data	76		Germany	20.5	**.248**[c]
Mayew, 2013, supplementary data	125		Japan	28.38	**.179**[d]
Loehr & Hara, 2013, from Dryad dataset, subset of sample	60		Finland	27.18	.232
Skorska et al., 2015	204		Canada	25.97	**.387**
Skorska et al., 2015		186	Canada	22.62	**.412**
Lefevre et al., 2014	54	49	UK	21.59	.090
Walker, Goetz, & Carré, 2015	146		USA	20.64	**.40**

Notes: Effect sizes in bold were reported as significant in the corresponding manuscripts. UK = United Kingdom. USA = United States of America. [a]positive values indicate that individuals with larger FWHRs have higher BMI than individuals with lower FWHRs. [b]These authors also provide results based on a meta-analysis of all of the effect sizes reported in their manuscript. Here, we only include the individual effect sizes. [c]Because there was some ambiguity in the number of participants in the analysis reported in the paper, we used the author's supplementary data set included with their manuscript to calculate this effect size. [d]The data sheet from which we derived this effect size had two entries for some players (information on BMI for two seasons). We only used data for the first season for such players.

References for Chapter 5 Supplementary Tables

Alrajih, S., & Ward, J. (2014). Increased facial width-to-height ratio and perceived dominance in the faces of the UK's leading business leaders. *British Journal of Psychology*, *105*(2), 153–61. http://doi.org/10.1111/bjop.12035

Anderson, C., & Galinsky, A. D. (2006). Power, optimism, and risk-taking. *European Journal of Social Psychology*, *36*(4), 511–536. http://doi.org/10.1002/ejsp.324

Bashir, N. Y., & Rule, N. O. (2014). Shopping under the influence: nonverbal appearance-based communicator cues affect consumer judgments. *Psychology of Marketing*, *31*, 539–548. http://doi.org/10.1002/mar

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*, 1–9. http://doi.org/10.1093/geronb/gbt054

Burton, C. M., & Rule, N. O. (2013). Judgments of height from faces are informed by dominance and facial maturity. *Social Cognition*, *31*(6), 672–685. http://doi.org/10.1521/soco.2013.31.6.672

Buss, A. H., & Perry, M. (1992). PERSONALITY PROCESSES AND INDIVIDUAL The Aggression Questionnaire, *63*(3), 452–459.

Buss, A. H., & Warren, W. L. (2000). *The aggression questionnaire manual*. Los Angeles, CA: Western Psychological Services.

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 of the Royal Society B Biological Sciences, *275*(1651), 2651–6. http://doi.org/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*(10), 1194–8. http://doi.org/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. http://doi.org/10.1068/p6543

Carré, J. M., Murphy, K. R., & Hariri, A. R. (2013). What lies beneath the face of aggression? *Social Cognitive and Affective Neuroscience*, *8*(2), 224–229. http://doi.org/10.1093/scan/nsr096

Cheng, J. T., Tracy, J. L., & Henrich, J. (2010). Pride, personality, and the evolutionary foundations of human social status. *Evolution and Human Behavior*, *31*, 334–347. http://doi.org/10.1016/j.evolhumbehav.2010.02.004.

Cherek, D. R. (1981). Psychopharmacology effects of smoking different doses of nicotine on human aggressive behavior. *Psychopharmacology*, *75*, 339–345.

Coetzee, V., Chen, J., Perrett, D. I., & Stephen, I. D. (2010). Deciphering faces: quantifiable visual cues to weight. *Perception*, *39*(1), 51–61. http://doi.org/10.1068/p6560

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*(2), 235–238. http://doi.org/10.1016/j.jrp.2012.01.005

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

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. http://doi.org/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*(1), e30366. http://doi.org/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*(5), 1027–43. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24192545

Geniole, S. N., & McCormick, C. M. (2015). Facing our ancestors: Judgements of aggression are consistent and related to the facial width-to-height ratio in men irrespective of beards. *Evolution and Human Behavior*, *36*(4), 279–285. http://doi.org/10.1016/j.evolhumbehav.2014.12.005

Geniole, S. N., Molnar, D. S., Carré, J. M., & McCormick, C. M. (2014). The facial width-to-height ratio shares stronger links with judgments of aggression than with judgments of trustworthiness. *Journal of Experimental Psychology. Human Perception and Performance*, *40*(4), 1526–1541. http://doi.org/10.1037/a0036732

Goetz, S. M. 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*(11), 2329–34. http://doi.org/10.1177/0956797613493294

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. J. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe* (pp. 7–28). Tillburg: Tilburg University Press.

Goldberg, L. R., Johnson, J. a., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*, 84–96. http://doi.org/10.1016/j.jrp.2005.08.007

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*(1), e52317. http://doi.org/10.1371/journal.pone.0052317

Haselhuhn, M. P., & Wong, E. M. (2012). Bad to the bone: facial structure predicts unethical behaviour. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 571–576. http://doi.org/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*(8), e72259. http://doi.org/10.1371/journal.pone.0072259

Haselhuhn, M. P., Wong, E. M., Ormiston, M. E., Inesi, M. E., & Galinsky, A. D. (2014). Negotiating face-to-face: Men’s facial structure predicts negotiation performance. *Leadership Quarterly*, *25*(5), 835–845. http://doi.org/10.1016/j.leaqua.2013.12.003

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*(3), 289–96. http://doi.org/10.1177/0956797612451467

Hehman, E., Leitner, J. B., & Freeman, J. B. (2014). The Face – Time Continuum : Lifespan Changes in Facial Width-to-Height Ratio Impact Aging-Associated Perceptions. *Personality and Social Psychology Bulletin*, *40*, 1624–1636. http://doi.org/10.1177/0146167214552791

Hehman, E., Leitner, J. B., & Gaertner, S. L. (2013). Enhancing static facial features increases intimidation. *Journal of Experimental Social Psychology*, *49*(4), 747–754. http://doi.org/10.1016/j.jesp.2013.02.015

Huh, H. (2013). Digit ratios, but not facial width-to-height ratios, are associated with the priority placed on attending to faces versus bodies. *Personality and Individual Differences*, *54*(1), 133–136. http://doi.org/10.1016/j.paid.2012.08.032

Huh, H., Yi, D., & Zhu, H. (2014). Facial width-to-height ratio and celebrity endorsements. *Personality and Individual Differences*, *68*, 43–47. http://doi.org/10.1016/j.paid.2014.03.044

Jia, Y., van Lent, L., & Zeng, Y. (2014). Masculinity, Testosterone, and Financial Misreporting. *Journal of Accounting Research*, *52*(5), 1195–1246. http://doi.org/10.1111/1475-679X.12065

Karp, R., Gold, D., & Tan, M. (2006). *Bullard Houses*. Cambridge, MA: Harvard University, Harvard Law School, Program on Negotiation.

Kleisner, K., Priplatova, L., Frost, P., & Flegr, J. (2013). Trustworthy-Looking Face Meets Brown Eyes. *PLoS ONE*, *8*(1), 1–7. http://doi.org/10.1371/journal.pone.0053285

Kramer, R. S. S., Jones, A. L., & Ward, R. (2012). A lack of sexual dimorphism in width-to-height ratio in white European faces using 2D photographs, 3D scans, and anthropometry. *PloS One*, *7*(8), e42705. http://doi.org/10.1371/journal.pone.0042705

Lefevre, C. E., Etchells, P. J., Howell, E. C., Clark, A. P., & Penton-voak, I. S. (2014). Facial width-to-height ratio predicts self-reported dominance and aggression in males and females , but a measure of masculinity does not. *Biology Letters*, *10*, 20140729. http://doi.org/10.1098/rsbl.2014.0729

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, (August). http://doi.org/10.1002/per

Lefevre, C. E., Lewis, G. J., Bates, T. C., Dzhelyova, M., Coetzee, V., Deary, I. J., & Perrett, D. I. (2012). No evidence for sexual dimorphism of facial width-to-height ratio in four large adult samples. *Evolution and Human Behavior*, *33*(6), 623–627. http://doi.org/10.1016/j.evolhumbehav.2012.03.002

Lefevre, C. E., Lewis, G. J., Perrett, D. I., & Penke, L. (2013). Telling facial metrics:

facial width is associated with testosterone levels in men. *Evolution and Human Behavior*, *34*(4), 273–279. http://doi.org/10.1016/j.evolhumbehav.2013.03.005

Lewis, G. J. J., Lefevre, C. E., & Bates, T. C. C. (2012). Facial width-to-height ratio predicts achievement drive in US presidents. *Personality and Individual Differences*, *52*(7), 855–857. http://doi.org/10.1016/j.paid.2011.12.030

Lilienfeld, S. O., & Widows, M. R. (2005). *Pychopathic Personality Inventory- Revised: Professional Manual*. Lutz, FL: Psychological Assessment Resources.

Loehr, J., & Hara, R. B. O. (2013). Facial morphology predicts male fitness and rank but not survival in Second World War Finnish soldiers Facial morphology predicts male fitness and rank but not survival in Second World War Finnish soldiers, (May).

Marsh, A. A., Cardinale, E. M., Chentsova-Dutton, Y. E., Grossman, M. R., & Krumpos, K. A. (2014). Power plays: Expressive mimicry of valid agonistic cues. *Social Psychological and Personality Science*, *5*(6), 684–690. http://doi.org/10.1177/1948550613519684

Mayew, W. J. (2013). Reassessing the association between facial structure and baseball performance: a comment on Tsujimura & Banissy (2013). *Biology Letters*, *9*, 20130538. http://doi.org/10.1098/rsbl.2013.0538

Mayew, W. J., Parsons, C. a., & Venkatachalam, M. (2013). Voice pitch and the labor market success of male chief executive officers. *Evolution and Human Behavior*, *34*(4), 243–248. http://doi.org/10.1016/j.evolhumbehav.2013.03.001

Mileva, V. R., Cowan, M. L., Cobey, K. D., Knowles, K. K., & Little, a. C. (2014). In the face of dominance: Self-perceived and other-perceived dominance are positively associated with facial-width-to-height ratio in men. *Personality and Individual Differences*, *69*, 115–118. http://doi.org/10.1016/j.paid.2014.05.019

Ostendorf, F. (2001). Measuring interpersonal behavior with the German Interpersonal Adjective Scales. In R. Riemann, F. M. Spinath, & F. Ostendorf (Eds.), *Personality and temperament: Genetics, evolution, and structure* (pp. 232–260). Lengerich, Germany: Pabst.

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*(3), 169–173. http://doi.org/10.1016/j.evolhumbehav.2011.08.001

Re, D. E., Hunter, D. W., Coetzee, V., Tiddeman, B. P., Xiao, D., DeBruine, L. M., … Perrett, D. I. (2013). Looking like a leader-facial shape predicts perceived height and leadership ability. *PLoS ONE*, *8*(12). http://doi.org/10.1371/journal.pone.0080957

Sanchez-pages, S., Rodriguez-ruiz, C., & Turiegano, E. (2014). Facial Masculinity : How the Choice of Measurement Method Enables to Detect Its Influence on Behaviour. *PLOS ONE*, *9*(11), e112157. http://doi.org/10.1371/journal.pone.0112157

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 : Official Journal of the Human Behavior and Evolution Society*, *33*(2), 121–129. http://doi.org/10.1016/j.evolhumbehav.2011.07.002

Skorska, M. N., Geniole, S. N., Vrysen, B. M., McCormick, C. M., & Bogaert, A. F. (2015). Facial Structure Predicts Sexual Orientation in Both Men and Women.

Archives of Sexual Behavior. http://doi.org/10.1007/s10508-014-0454-4

Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: male facial width and trustworthiness. *Psychological Science*, *21*(3), 349–54. http://doi.org/10.1177/0956797610362647

Stirrat, M., & Perrett, D. I. (2012). Face Structure Predicts Cooperation: Men With Wider Faces Are More Generous to Their In-Group When Out-Group Competition Is Salient. *Psychological Science*, *23*(7), 718–22. http://doi.org/10.1177/0956797611435133

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*(5), 551–556. http://doi.org/10.1016/j.evolhumbehav.2012.02.002

Třebický, V., Fialová, J., Kleisner, K., Roberts, S. C., Little, A. C., & Havlíček, J. (2015). Further Evidence for Links Between Facial Width-to-Height Ratio and Fighting Success : Commentary on Zilioli et al.(2014). *Aggressive Behavior*, *41*, 331–334. http://doi.org/10.1002/ab.21559

Tsujimura, H., & Banissy, M. J. (2013). Human face structure correlates with professional baseball performance: insights from professional Japanese baseball players. *Biology Letters*, *9*, 20130140. http://doi.org/10.1098/rsbl.2013.0140

Valentine, K. a., Li, N. P., Penke, L., & Perrett, D. I. (2014). Judging a man by the width of his face: the role of facial ratios and dominance in mate choice at speed-dating events. *Psychological Science*, *25*(3), 806–11.

http://doi.org/10.1177/0956797613511823

Wang, Y., Geigel, J., & Herbert, A. (2013). Reading personality: Avatar vs. human faces. *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*, 479–484. http://doi.org/10.1109/ACII.2013.85

Welker, K. M., Goetz, S. M. M., & Carré, J. M. (2015). Perceived and experimentally manipulated status moderate the relationship between facial structure and risk-taking. *Evolution and Human Behavior*. http://doi.org/10.1016/j.evolhumbehav.2015.03.006

Welker, K. M., Goetz, S. M. M., Galicia, S., Liphardt, J., & Carré, J. M. (2014). An Examination of the Associations Between Facial Structure, Aggressive Behavior, and Performance in the 2010 World Cup Association Football Players. *Adaptive Human Behavior and Physiology*. http://doi.org/10.1007/s40750-014-0003-3

Weston, E. M., Friday, A. E., & Liò, P. (2007). Biometric evidence that sexual selection has shaped the hominin face. *PloS One*, *2*(8), e710. http://doi.org/10.1371/journal.pone.0000710

Wong, E. M., Ormiston, M. E., & Haselhuhn, M. P. (2011). A face only an investor could love: CEOs’ facial structure predicts their firms' financial performance. *Psychological Science*, *22*(12), 1478–83. http://doi.org/10.1177/0956797611418838

Zilioli, S., Sell, A. N., Stirrat, M., Jagore, J., Vickerman, W., & Watson, N. V. (2015). Face of a fighter: Bizygomatic width as a cue of formidability. *Aggressive Behavior*, *41*, 322–330. http://doi.org/10.1002/ab.21544

Chapter 5 Supplementary Figures

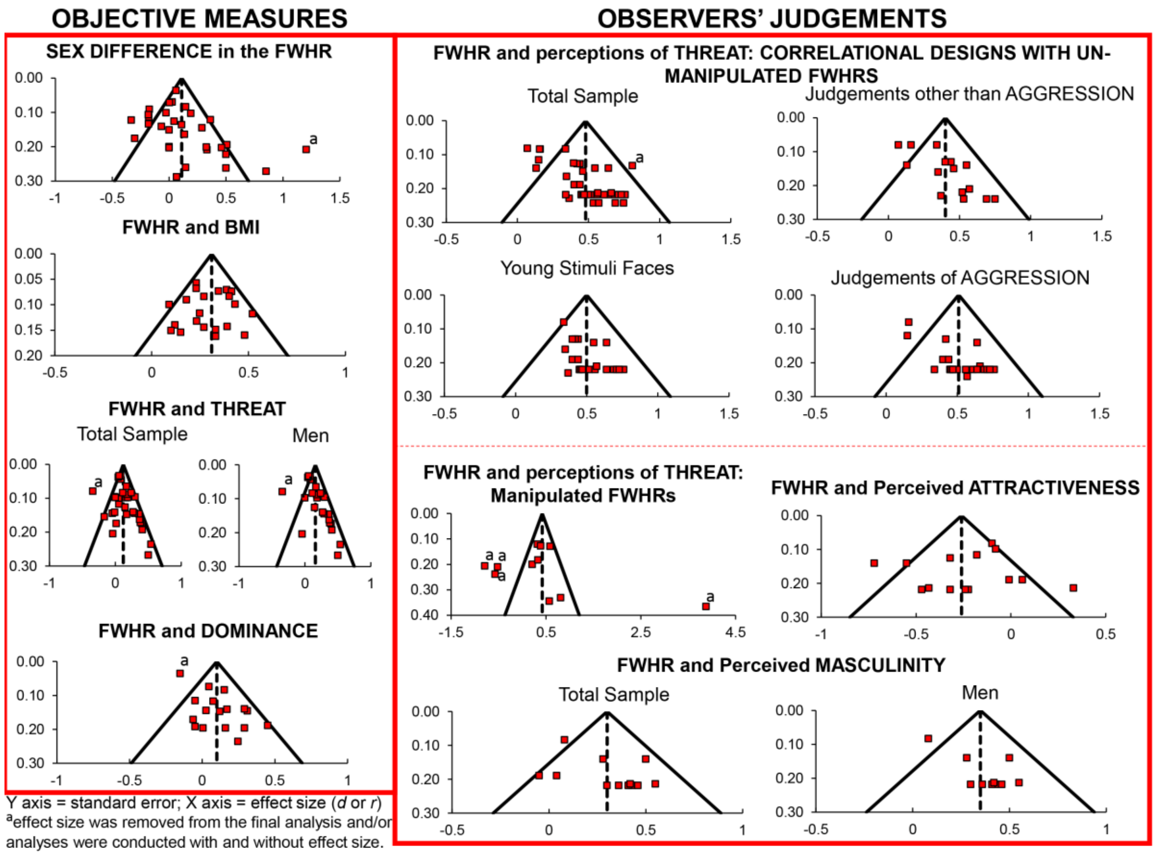

Figure S5-1. Funnel plots of effect sizes.

Appendix A

Brock University
Research Ethics Board
Tel: 905-688-5550 ext. 3035
Email: reb@brocku.ca

Certificate of Ethics Clearance for Human Participant Research

DATE:	11/17/2010		
PRINCIPAL INVESTIGATOR:	MCCORMICK, Cheryl - Psychology		
FILE:	10-087 - MCCORMICK		
TYPE:	Masters Thesis/Project	STUDENT: SUPERVISOR:	Shawn Geniole Cheryl McCormick

TITLE: Relationship between personality, salivary hormones and strategic decision-making

ETHICS CLEARANCE GRANTED

Type of Clearance: NEW Expiry Date: 11/30/2011

The Brock University Research Ethics Board has reviewed the above named research proposal and considers the procedures, as described by the applicant, to conform to the University's ethical standards and the Tri-Council Policy Statement. Clearance granted from **11/17/2010** to **11/30/2011**.

The Tri-Council Policy Statement requires that ongoing research be monitored by, at a minimum, an annual report. Should your project extend beyond the expiry date, you are required to submit a Renewal form before **11/30/2011**. Continued clearance is contingent on timely submission of reports.

To comply with the Tri-Council Policy Statement, you must also submit a final report upon completion of your project. All report forms can be found on the Research Ethics web page.

In addition, throughout your research, you must report promptly to the REB:

a) Changes increasing the risk to the participant(s) and/or affecting significantly the conduct of the study;
b) All adverse and/or unanticipated experiences or events that may have real or potential unfavourable implications for participants;
c) New information that may adversely affect the safety of the participants or the conduct of the study;
d) Any changes in your source of funding or new funding to a previously unfunded project.

We wish you success with your research.

Approved: Michelle K. McC

Michelle McGinn, Chair
Research Ethics Board (REB)

<u>Note:</u> Brock University is accountable for the research carried out in its own jurisdiction or under its auspices and may refuse certain research even though the REB has found it ethically acceptable.

If research participants are in the care of a health facility, at a school, or other institution or community organization, it is the responsibility of the Principal Investigator to ensure that the ethical guidelines and clearance of those facilities or institutions are obtained and filed with the REB prior to the initiation of research at that site.

Appendix B

Brock University
Research Ethics Office
Tel: 905-688-5550 ext. 3035
Email: reb@brocku.ca

Bioscience Research Ethics Board

Certificate of Ethics Clearance for Human Participant Research

DATE:	October 26, 2011		
PRINCIPAL INVESTIGATOR:	MCCORMICK, Cheryl - Psychology		
FILE:	10-087 - MCCORMICK		
TYPE:	Masters Thesis/Project	STUDENT: SUPERVISOR:	ShawnGeniole Cheryl McCormick

TITLE: Relationship between personality, salivary hormones and strategic decision-making

ETHICS CLEARANCE GRANTED

Type of Clearance: MODIFICATION Expiry Date: 9/28/2012

The Brock University Bioscience Research Ethics Board has reviewed the above named research proposal and considers the procedures, as described by the applicant, to conform to the University's ethical standards and the Tri-Council Policy Statement. Clearance granted from **10/26/2011** to **9/28/2012**.

The Tri-Council Policy Statement requires that ongoing research be monitored by, at a minimum, an annual report. Should your project extend beyond the expiry date, you are required to submit a Renewal form before **9/28/2012**. Continued clearance is contingent on timely submission of reports.

To comply with the Tri-Council Policy Statement, you must also submit a final report upon completion of your project. All report forms can be found on the Research Ethics web page at http://www.brocku.ca/research/policies-and-forms/research-forms.

In addition, throughout your research, you must report promptly to the REB:

a) Changes increasing the risk to the participant(s) and/or affecting significantly the conduct of the study;
b) All adverse and/or unanticipated experiences or events that may have real or potential unfavourable implications for participants;
c) New information that may adversely affect the safety of the participants or the conduct of the study;
d) Any changes in your source of funding or new funding to a previously unfunded project.

We wish you success with your research.

Approved:

Brian Roy, Chair
Bioscience Research Ethics Board

Note: Brock University is accountable for the research carried out in its own jurisdiction or under its auspices and may refuse certain research even though the REB has found it ethically acceptable.

If research participants are in the care of a health facility, at a school, or other institution or community organization, it is the responsibility of the Principal Investigator to ensure that the ethical guidelines and clearance of those facilities or institutions are obtained and filed with the REB prior to the initiation of research at that site.

Appendix C

Brock University
Research Ethics Office
Tel: 905-688-5550 ext. 3035
Email: reb@brocku.ca

Bioscience Research Ethics Board

Certificate of Ethics Clearance for Human Participant Research

DATE: November 15, 2013

PRINCIPAL INVESTIGATOR: MCCORMICK, Cheryl - Psychology
FILE: 10-087 - MCCORMICK

TYPE: Ph. D. STUDENT: Shawn Geniole
SUPERVISOR: Cheryl McCormick

TITLE: Relationship between personality, salivary hormones and strategic decision-making

ETHICS CLEARANCE GRANTED

Type of Clearance: MODIFICATION Expiry Date: 9/30/2014

The Brock University Bioscience Research Ethics Board has reviewed the above named research proposal and considers the procedures, as described by the applicant, to conform to the University's ethical standards and the Tri-Council Policy Statement. Clearance granted from **11/15/2013** to **9/30/2014**.

The Tri-Council Policy Statement requires that ongoing research be monitored by, at a minimum, an annual report. Should your project extend beyond the expiry date, you are required to submit a Renewal form before **9/30/2014**. Continued clearance is contingent on timely submission of reports.

To comply with the Tri-Council Policy Statement, you must also submit a final report upon completion of your project. All report forms can be found on the Research Ethics web page at http://www.brocku.ca/research/policies-and-forms/research-forms.

In addition, throughout your research, you must report promptly to the REB:

a) Changes increasing the risk to the participant(s) and/or affecting significantly the conduct of the study;
b) All adverse and/or unanticipated experiences or events that may have real or potential unfavourable implications for participants;
c) New information that may adversely affect the safety of the participants or the conduct of the study;
d) Any changes in your source of funding or new funding to a previously unfunded project.

We wish you success with your research.

Approved:

Brian Roy, Chair
Bioscience Research Ethics Board

Note: Brock University is accountable for the research carried out in its own jurisdiction or under its auspices and may refuse certain research even though the REB has found it ethically acceptable.

If research participants are in the care of a health facility, at a school, or other institution or community organization, it is the responsibility of the Principal Investigator to ensure that the ethical guidelines and clearance of those facilities or institutions are obtained and filed with the REB prior to the initiation of research at that site.