Justice is (not so) blind: Effects of facial masculinity and agreeableness on perceptions of criminal guilt.

Kathryn D. Ford[1], Ian S. Penton-Voak[2] & Nicholas Pound[3]

[1] Centre for Culture & Evolution, Department of Life Sciences, Brunel University London.
Email: kathryn.ford@brunel.ac.uk

[2] School of Psychological Science, University of Bristol.
Email: i.s.penton-voak@bristol.ac.uk

[3] Centre for Culture & Evolution, Department of Life Sciences, Brunel University London.
Email: nicholas.pound@brunel.ac.uk

Correspondence:

Dr Nicholas Pound
Centre for Culture & Evolution
Department of Life Sciences
College of Health, Medicine & Life Sciences
Brunel University London
Uxbridge, UB8 3PH
United Kingdom.
Email: nicholas.pound@brunel.ac.uk

Abstract

People have a tendency to make rapid judgements about the personality of others based on their facial appearance, a tendency which could have adaptive value if it helps facilitate the avoidance of individuals disposed to exploit and/or harm the perceiver. These rapid judgements, accurate or not, have the potential to influence how individuals are treated in many areas of life, including within the criminal justice system. Previous research investigating effects of appearance on judicial proceedings has suggested that a masculine facial appearance might activate criminal stereotypes, and therefore increase the likelihood of being judged guilty of a crime. To examine how masculinity might interact with other appearance dimensions, we investigated how facial morphological masculinity and perceived agreeableness influence perceptions of criminal guilt. In an online study, 369 participants (167 men, 200 women, 2 did not say) aged 18 to 82, read 12 short vignettes each describing a fictional crime (assault, burglary or rape) with each accompanied by the face of a man "charged" with the crime. Faces were manipulated using morphing techniques to increase or decrease levels of a) morphological masculinity and b) perceived agreeableness (i.e. 2 x 2 manipulations for each target face). Participants were asked to indicate in each case whether they thought the "accused" was guilty or not. Overall, facial appearance had a significant effect on the probability of being judged guilty. For each crime type, manipulations of perceived agreeableness had large effects on the probability of being judged guilty, whereas manipulations of morphological masculinity did not.

Public Significance Statement

This study investigated how facial appearance can affect whether people are perceived as likely to have committed a crime. Computer graphics techniques were used to manipulate aspects of male facial shape and the facial images were presented to volunteer participants in an online study. Manipulations of perceived agreeableness in faces had large effects on the probability of an individual being judged guilty of a hypothetical crime. However, changes to the shape of the face to make it more or less masculine did not.

Introduction

The idea that it is possible to assess someone's personality from their face dates back thousands of years (Liggett, 1974). There is some limited evidence that a 'kernel of truth' may underpin some of these judgements (for a review see Zebrowitz & Collins, 1997). However, levels of accuracy are generally quite low (e.g. Penton-Voak, Pound, Little, Perrett, 2006). Nevertheless, the idea that the face provides a valid guide to personality and behaviour remains popular (e.g. Hassin & Trope, 2000) and people do tend to automatically make rapid judgements about the personality and behaviour of others based on facial appearance alone (e.g. Willis & Todorov, 2006; for a review see Toderov, Said & Verosky, 2012). Moreover, researchers adopting an evolutionary perspective have suggested that this tendency could reflect of the functioning of specialised learning mechanisms that have an important role in interpersonal perception (e.g. Haselton & Funder, 2006).

An ecological approach to social perception suggests that humans have evolved perceptual skills (or "attunements") that serve an adaptive function in guiding social behaviour (McArthur & Baron, 1983; Zebrowitz & Collins, 1997). Specifically, the ability to make accurate predictions about the behavioural dispositions of others can help individuals to "navigate" social relationships (Huelsnitz, Neel & Human, 2020). That is, accurate perceptions of personality in others can help inform decisions about which conspecifics to engage with cooperatively, and which to endeavour to avoid (e.g. Mayer, Phillips and Barry, 2015; Sacco & Brown, 2018). Such perceptions would allow individuals to preferentially interact with social partners who possess personality attributes that complement their own personality (e.g. Brown, Sacco & Medlin, 2019) or are likely to be of particular value in specific environments (Sacco & Brown, 2018). Moreover, they would facilitate the avoidance of individuals who are disposed to exploit and/or harm the perceiver, such as those displaying

facial cues to traits such as untrustworthiness (Stirrat & Perret, 2010), narcissism (Medlin, Sacco & Brown, 2020) and psychopathy (Brown, Sacco, Lolley & Block, 2017).

Character judgements based on facial appearance.

Perhaps the most widely studied example of how people make rapid judgements about the personality and behaviour of others based on facial appearance is the attractiveness "halo" effect, whereby attractive individuals are judged more positively than less attractive individuals on a range of socially desirable dimensions (e.g. Dion, Berscheid, & Walster, 1972; Zebrowitz, Voinescu, & Collins, 1997). Another widely studied, and related, example is the "babyface" overgeneralisation effect – i.e. the manner in which individuals with childlike facial features are judged to possess childlike behavioural traits such as honesty and trustworthiness (e.g. Berry & McArthur, 1985; Zebrowitz & Montepare, 1992; Zebrowitz et al., 2012; Zebrowitz & Franklin, 2014). According to the ecological approach, tendencies such as these are best thought of as arising from "overgeneralizations of highly adaptive perceptual attunements" (McArthur & Baron, 1983, p. 231) and accurate or not, character judgements such as these, based on facial appearance alone, have the potential to influence how people are treated in many areas of life.

Examples of the influence such judgements can exert come from studies showing that facial attractiveness can increase the success of job applications accompanied by photographs (Maurer-Fazio & Lei, 2015) and judgements of competence based on facial appearance can predict the success of candidates in elections (e.g. Todorov et al., 2005; Marcinkowski, Lünich, & Starke, 2018). Also, importantly, facial appearance has the potential to affect how individuals are treated by the criminal justice system – a domain in which attributions of dishonesty and untrustworthiness could have major implications. For example, using photographs of men convicted of murder Wilson & Rule (2015) showed that perceptions of trustworthiness based on facial appearance predicted severity of sentencing. Moreover, this

sentencing bias was even found in a sample of men who were later exonerated, suggesting that the finding was not due a real association between facial appearance and the specific nature of the crimes committed. However, as post-sentencing photos were used in that study, it is possible that a "Dorian Grey effect contributed to the findings.

Facial appearance and the criminal justice system

Studies within the legal system itself pose practical difficulties, and there is limited scope to control for potentially confounding factors in naturalistic stimuli. Consequently, experimental investigations of extra-legal factors that may influence judgements of criminal guilt commonly use "mock jury" paradigms, where participants (either individually or in a group) are asked to respond with verdicts following the presentation of information about a scenario - usually in the form of "vignettes" (Bieneck, 2015). Using such techniques, mock jury studies have shown that attractive individuals are less likely to be judged guilty of crimes (e.g. Darby & Jeffers, 1988; Efran, 1974; for a meta-analysis see Mazzella & Feingold, 1994) and when found guilty they receive less severe sentences (e.g. Darby & Jeffers, 1988; DeSantis & Kayson 1997; Leventhal and Krate, 1977). Moreover, in mock litigation cases, attractive plaintiffs are more likely to be successful and are awarded more money in damages (Kulka & Kessler 1978). In real courtrooms, there is evidence that both attractiveness and baby-facedness can increase the likelihood of success for individuals involved in small claims court litigation cases (Zebrowitz & McDonald 1991). Moreover, observational studies suggest that attractiveness can influence the severity of sentencing in real criminal cases (Stewart, 1980; 1985).

As noted above, there is evidence that the facial attractiveness "halo" effect and the "babyface" overgeneralisation effect have the potential to influence the outcome of legal proceedings. However, another facial attribute that could elicit negative judgements relevant to legal proceedings through overgeneralisation mechanisms is facial masculinity – as a

masculine appearance could activate "criminal stereotypes" (Ward, Flowe & Humphries, 2012). It is widely recognised that men are more likely than women to commit a range of crimes (Rowe, Flannery & Flannery, 1995; Wilson, & Herrnstein, 1985). Consequently, exaggerated male-like facial characteristics could elicit greater suspicion of criminal guilt, directly if involvement in crime is seen as a more male-typical behavioural trait, but also indirectly as men with more masculine faces are seen as more dominant, less cooperative, and less honest (Perrett, et al, 1998; Johnston et al., 2001). Whether or not personality attributions elicited by masculine facial features have any accuracy, they have the potential to influence the outcome of legal proceedings. For example, Ward et al. (2012) found that men rated as having a more masculine facial appearance were rated as being more likely to be guilty of various offences including burglary, fraud and forgery. Moreover, a recent study by Estrada-Reynolds et al (2017) found, using a mock line-up approach, that increasing levels of rated facial masculinity increased the likelihood of a target face being selected as the perpetrator of an assault, but did not influence the chances of being selected as the perpetrator of a burglary or fraud. This study used face morphing procedures to create faces with moderate levels of rated facial masculinity. However, as in Ward et al (2012), facial masculinity was ultimately defined according to subjective ratings of masculinity in the faces rather than objective anatomical measurements (or direct manipulation), of *morphological masculinity* (see below).

Studying the role of facial masculinity

The use of subjective ratings of facial masculinity to create stimuli is potentially problematic, as judgements of "masculinity" may not be entirely independent from perceptions of traits such as aggressiveness and dishonesty in the target faces, thereby introducing some circularity to the evidence. There may be "cross-contamination" (Penton-Voak et al., 2006) with judgements of one trait influencing judgements of another trait. For example, if perceptions of aggressiveness lead a face to be rated as more masculine, then it

should not be surprising that stimuli created from faces rated as masculine are then judged to be more likely to commit violent crimes.

Instead of relying on masculinity ratings, an alternative approach is to measure *morphological masculinity* in faces objectively (see for a review see Mitteroecker et al., 2015) and/or use computer graphics face transforming techniques (for a review Sutherland, Rhodes & Young, 2017) to create stimuli that vary in objectively measured facial morphological masculinity. Human faces exhibit sexual dimorphism along various dimensions – i.e. there are shape differences between the average male and average female face and individual faces lie along a continuum defined by the differences between these. Men with face shapes close to the female average can be described as having a low level of morphological masculinity, while moving away from the female average towards (or beyond) the male average means a face exhibits a higher level of morphological masculinity. Importantly, subjective ratings of "masculinity" are often only moderately correlated with objective measures of morphological masculinity (e.g. Mitteroecker et al., 2015; Sanchez-Pages, Rodriguez-Ruiz, & Turiegano, 2014).

The sexually dimorphic continuum described above can be used as a basis for creating standardised face stimuli that exhibit high vs low levels of morphological masculinity. This is achieved using average male and average female prototypes as anchors, and employing computer graphics face transforming techniques to increase or decrease levels of morphological masculinity relative to these anchors – i.e. move them along the masculine-feminine continuum (e.g. DeBruine, Jones, Crawford, Welling, & Little, 2010; Scott et al., 2014). This technique has been used in the present study to establish whether direct manipulations of morphological facial masculinity can influence the extent to which an individual is perceived as likely to have committed a crime.

Two-dimensional structure of face evaluations

Facial masculinity is an important contributor to judgements of "dominance" – i.e. the *ability* of an individual to exert power over others (Todorov et al., 2008). However, whether someone has the *intention* to cause harm is of course also an important consideration when making judgements about whether someone is likely to have committed a crime. Moreover, as noted earlier, accurate perception of this second dimension likely has particular adaptive value in that it enables the avoidance of individuals likely to exploit and/or harm the perceiver (Brown, Sacco, Lolley & Block, 2017). Indeed, McArthur & Baron's (1983) early formulation of the ecological theory of social perception proposed that the need to distinguish individuals with "benevolent" versus "malevolent" intent was one of the key adaptive challenges that human perceptual systems will have evolved to deal with. Moreover, data-driven research has shown that two dimensions, that arguably reflect the *ability* and *intention* to cause harm respectively, are central to how people evaluate the faces of others. When people are asked to make judgements about people on a range of traits based on facial appearance alone, principal components analysis has shown that these load on two fundamental dimensions that correspond closely to judgements of power/dominance and valence/trustworthiness (Oosterhof & Todorov, 2008; Todorov et al., 2008).

Flowe (2012) examined whether this 2-dimensional (2D) model of face evaluation can help explain why some individuals are perceived as more likely to have committed crimes. That study showed, using naturalistic photographs, that perceived dominance was strongly positively correlated, and perceived trustworthiness, was strongly negatively correlated, with perceived criminality. Moreover, in accordance with the 2D model, these two dimensions could account for the majority of the variability in the criminality ratings. However, the use of naturalistic photographs meant that the underlying aspects of facial appearance that contribute to judgements of dominance and trustworthiness were not manipulated experimentally.

9

The present study

In the present study, as in Flowe (2012) we sought apply the 2D model (Oosterhof & Todorov, 2008; Todorov et al., 2008) but using experimental manipulations of facial appearance – to examine the effects of each dimension on judgements about whether someone is likely to have committed a crime. Facial morphological masculinity (arguably a predictor of the *ability* to cause harm) was manipulated using the computer graphics face transforming technique outlined above. In addition, a similar technique was used to manipulate cues to the second dimension (i.e. the *intention* to cause harm). Here, the second dimension was operationalised as being perceived as scoring low on the "Big Five" personality dimension of agreeableness, a construct that is generally taken to encompass traits such as being prosocial, cooperative, trusting, considerate and kind (for reviews of the "Big Five" taxonomy see Digman, 1990; John & Srivastava, 1999).

Face stimuli that vary on a dimension of perceived personality (e.g. perceived agreeableness) can be created by using as anchors for the transform process, prototype faces that have been created by averaging groups of faces rated high vs low on that particular perceived dimension (Sutherland, Rhodes & Young, 2017). Accordingly, here groups of faces rated as high or low on agreeableness were used to manipulate cues to the second dimension (i.e. the *intention* to cause harm).To summarise, the present study was conducted to establish whether direct manipulations of morphological facial masculinity can influence judgements of guilt for certain criminal offences (assault, burglary and rape). In addition, in accordance with the 2D model of face evaluation (Oosterhof & Todorov, 2008; Todorov et al., 2008) applied by Flowe (2012) in a similar context, we also directly manipulated perceived agreeableness in the faces to examine whether judgements of guilt might depend on interactions between cues to the *ability* to harm (masculinity), and cues to the *intention* to harm).. It was hypothesised

that high levels of facial morphological masculinity and low levels of perceived agreeableness would both increase the likelihood of being judged guilty of a crime.

Methods

Participants

Participants were initially 486 volunteers aged 18 to 82 ($M = 38.8$, $SD = 16.8$) recruited via advertisements posted to UK community social media sites and directed to an online survey hosted by SurveyMonkey.com (240 men, 241 women and 5 who did not indicate a gender). However, as explained in the results section below, analyses presented here are based on data from the 369 participants who completed all items and did not respond in an identical manner to all items. These participants (167 men, 200 women and 2 who did not indicate a gender) were aged 18 to 82 ($M = 38.6$, $SD = 16.9$). The study was approved by the Brunel University Department of Psychology Research Ethics committee.

Materials

Vignettes

Information about fictitious crimes was presented in the form of 12 short vignettes written in the style of brief police press releases. In this way, the study attempted to replicate the rapid judgements about the likelihood of guilt that people might make after seeing reports about crimes in newspapers, on television, or on news websites, rather than in a courtroom where more information would be available. There were 4 vignettes for each of 3 types of crime (assault, burglary, rape). Presenting the vignettes in the form of brief press releases ensured that the vignettes contained no potentially distressing graphic details yet presented information in a concise and plausible manner, similar to announcements that might be made on television or radio news bulletins, or in newspapers. For example, a vignette concerning an assault took the form "*Police investigating an assault which happened on the evening of 25th*

11

January 2012 on Bold Street, Liverpool have charged John Doe, with Assault Occasioning Actual Bodily Harm. He will appear before Liverpool Magistrates' Court on 28th February 2012." Plausible fictitious names were used for suspects (but omitted here) and reports were dated some time prior to data collection so that participants could reasonably be asked whether they thought the individual concerned was subsequently convicted of the crime.

Face Stimuli

To avoid associating any particular identifiable individual with a crime, composite photographs were used as stimuli rather than photographs of individual men. To create these composites, digital photographs for 36 men aged 18-24 were used. All had been photographed as part of a previous research project. All previous research was ethically approved and all participants had given their informed consent for their photographs to be used for research purposes (in any case the composite creation process rendered them anonymous and non-identifiable in the present study). The 36 faces were split into 12 groups of 3 faces and for each face, the 219 facial landmark defined by Stephan et al (2005) were delineated using Psychomorph (Perception Lab, University of St Andrews). For each group, these landmarks and colour/texture information were used to create an "average" of the 3 faces. For details of the methods see Tiddeman, Burt, & Perrett (2001).

Face manipulations

The 12 male composite faces were then manipulated in order to vary their appearance along 2 dimensions 1) morphological masculinity vs femininity; and 2) perceived agreeableness vs disagreeableness. To achieve this, 2 x 2 manipulation morphologically masculinised and feminized versions of each face were first created using prototype-based computer graphics transformations (Tiddeman et al., 2001). The shape of each face was transformed relative to the differences between two prototype faces using Psychomorph

(Perception Lab, University of St Andrews). One prototype was an average male face (a composite of 20 male faces) and the other was an average female face (a composite of 20 female faces). For each stimulus face, the shape was transformed by ±50% of the shape difference between the two prototype faces which served as anchors to create a morphologically masculinised version and a feminized version. The technique used here to create face stimuli that vary along dimensions of sexual dimorphism has been very widely used since it was first employed by Perrett et al (1998) to investigate female variation in preferences for facial masculinity (for reviews see DeBruine et al, 2010; Sutherland, Rhodes, & Young, 2017). Moreover, it is well-established that the ±50% morphing technique in particular, reliably produces pairs of face stimuli that differ in perceived dominance and perceived masculinity (e.g. Alharbi et al., 2020; DeBruine et al., 2006; Perrett et al, 1998). For more details of the methods and computer algorithms involved in the face shape transformations see Rowland & Perrett (1995).

Following this, the 12 morphologically masculinized and 12 morphologically feminized versions of each face were then further manipulated to create agreeable and disagreeable versions of each using a similar prototype-based technique with the 2 anchors being agreeable vs disagreeable prototypes. The agreeable anchor was a composite of the 15 male faces rated as appearing most agreeable from a larger sample of 146 men in a previous study (Penton-Voak et al., 2006). The disagreeable anchor was a composite of the 15 male faces rated as appearing least agreeable from that study. In that study, agreeableness ratings were obtained on a 7-point scale anchored with an adjective pair (Suspicious-Trusting) that loaded highly on the agreeableness dimension in the factor analysis of self-report personality questionnaire data reported in the same study.

For each stimulus face, the shape was transformed by 50% of the shape difference between the two prototype faces. This process yielded 4 versions of each of the 12 composite

faces; a) feminized-agreeable, b) feminized-disagreeable, c) masculinized-agreeable, d) masculinized-disagreeable. All faces were then masked to remove hair and clothing cues. Figure 1 shows an example of this 2 x 2 manipulation, applied to a composite of 3 faces drawn from the publicly accessible Face Research Lab London Set (DeBruine & Jones, 2017).

Procedure

Participants read each of the 12 vignettes accompanied by one of the face stimuli with an indication that this was the face of the man "charged" with the crime described. For counterbalancing, 3 versions of the experiment were created so that each of the 12 faces appeared "accused" of the 3 crime types -Version A (Assault, Faces 1-4; Burglary, Faces 5-8, Rape, Faces 9-12), Version C (Assault, Faces 9-12; Burglary, Faces 1-4, Rape, Faces 5-8), Version C (Assault, Faces 5-8; Burglary, Faces 9-12; Rape, Faces 1-4). Due to the limitations of the randomization options offered by the online survey system used, the 3 versions were presented to 3 sequential groups of participants ($n = 216$; $n = 120$; $n = 150$ in the initial sample; of these n = 178; n = 81; n = 110 respectively completed the study and are included in the analyses) to achieve counterbalancing. For each vignette, the manipulated version of the target face shown was randomly selected (i.e. a) feminized-agreeable, b) feminized-disagreeable, c) masculinized-agreeable, or d) masculinized-disagreeable). Participants were asked to indicate whether they though the "accused" was guilty or not guilty of the crime described. Then they were also asked to indicate how "confident" they were about their judgement on a 7-point scale from 1 (Not at all confident) to 7 (Very confident). Guilt judgement data were analysed using chi-squares with the proportion of trials on which faces were judged guilty of crimes as the dependent measure. Participant confidence judgements were analysed using repeated measures ANOVA (IBM SPSS 26.0).
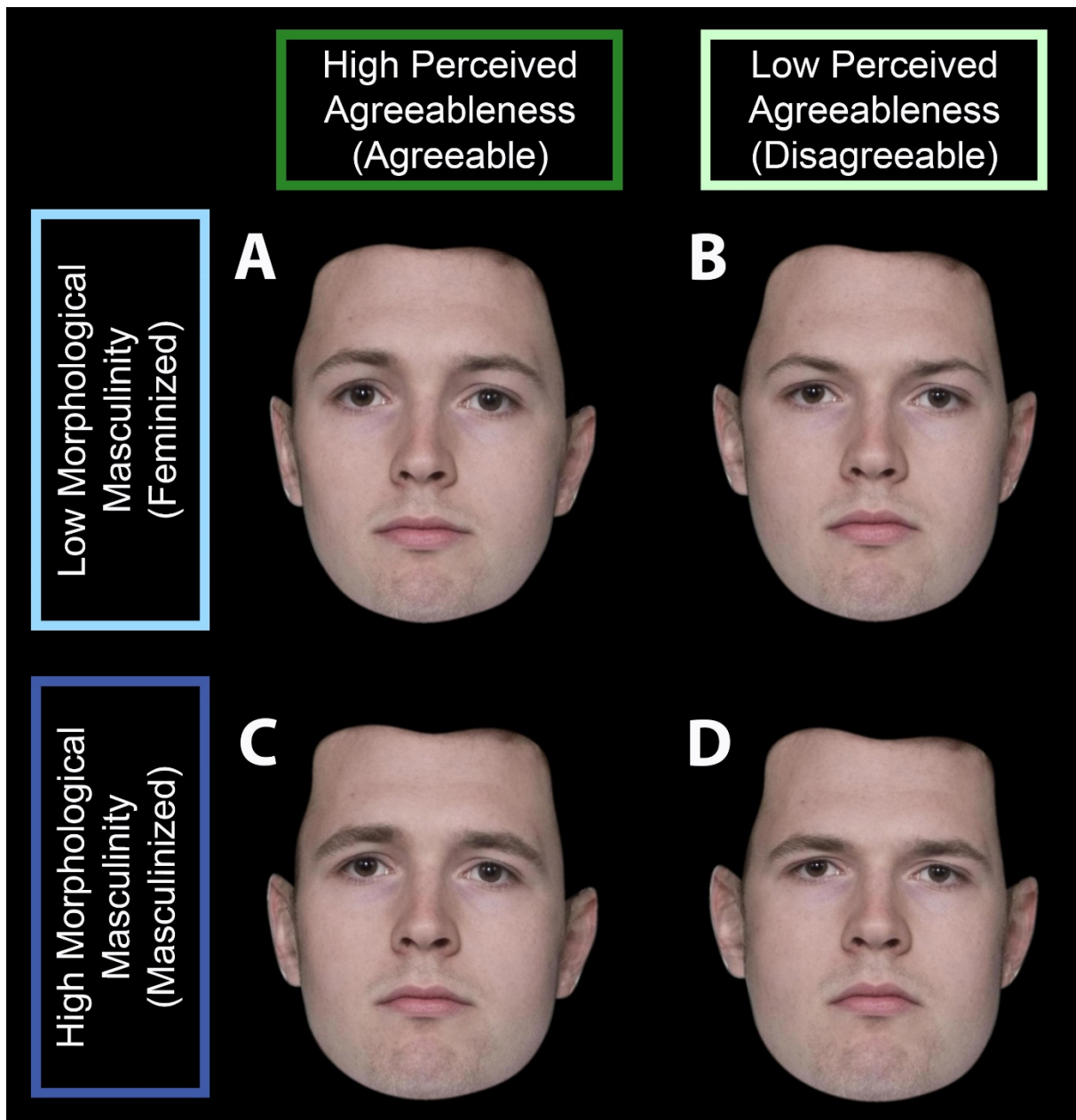
Figure 1: Four versions of a single composite face: a) feminized-agreeable, b) feminized-disagreeable, c) masculinized-agreeable, d) masculinized-disagreeable. Composite for this example (not used in present study) was derived from 3 individual faces in the Face Research Lab London Set (DeBruine & Jones, 2017).

Results

Of the 486 participants who started the study, only 416 provided responses to all 12 vignettes. The 26 participants who responded "not guilty" to all vignettes and the 21 who responded "guilty" to all vignettes were excluded and the analyses presented are based on data from the remaining 369 participants who did not respond in an identical manner to all items. Overall, participants judged the "accused" to be guilty in 58.1% of the 4428 trials and the proportion of guilty judgements did not differ across crime types; assault (57.4%), burglary (57.4%), rape (59.6%), $\chi^2 = 2.02$; $df = 2$; $p = .36$.

As shown in Table 1, collapsing across vignettes, treating the 4 face types as nominal categories (feminized-agreeable, feminized-disagreeable, masculinized-agreeable, masculinized-disagreeable), a series of 4 x 2 (Face Type x Verdict) chi-square tests (one for each crime type) showed that overall facial appearance had a significant effect on the probability of being judged guilty of assault ($\chi^2 = 90.33$; $df = 3$; $p < .001$), burglary ($\chi^2 = 10.35$; $df = 3$; $p = .016$) and rape ($\chi^2 = 36.08$; $df = 3$; $p < .001$).

Table 1: Percentage of each face type judged guilty for each crime type.

	Low Morphological Masculinity (Feminized)		High Morphological Masculinity (Masculinized)	
	a) Agreeable	b) Disagreeable	c) Agreeable	d) Disagreeable
Assault	42.7%	70.7%	47.8%	67.4%
Burglary	53.8%	61.1%	52.6%	61.8%
Rape	52.7%	67.2%	50.4%	66.4%

However, as shown in Figure 2, collapsing across agreeableness manipulations, a series of 2 x 2 (Masculinity x Verdict) chi-square tests (one for each crime type) showed that facial masculinity had no significant effects on perceptions of guilt for any crime type (all $\chi^2 <$ 0.80; $df = 1$; $p > .38$). In contrast, however, as shown in Figure 3 collapsing across masculinity manipulations, a series of 2 x 2 (Agreeableness x Verdict) chi-square (one for each crime type) showed that agreeableness had significant effects on perceptions of guilt for assault ($\chi^2 = 87.60$; $df = 1$; $p < .001$), burglary ($\chi^2 = 10.20$; $df = 1$; $p = .0014$) and rape ($\chi^2 = 35.65$; $df = 1$; $p < .001$). Repeating the analysis separately for male and female participants, similar patterns were found (See Supplemental Material).
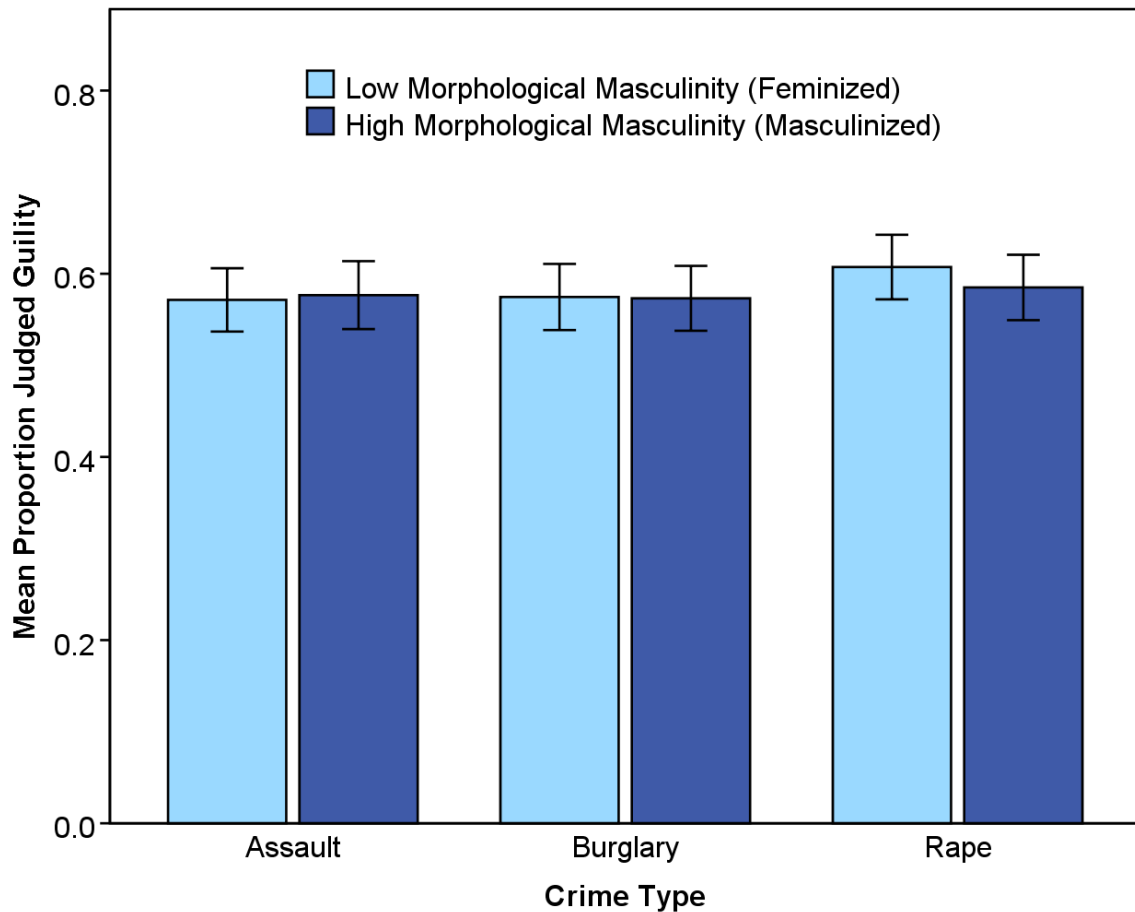
Figure 2: Effects of facial morphological masculinity manipulations on probability of being judged guilty of a crime, collapsed across agreeableness manipulations (error bars = 95% CI). Aggregate data for 4428 judgements made by 369 participants viewing 12 photos each.
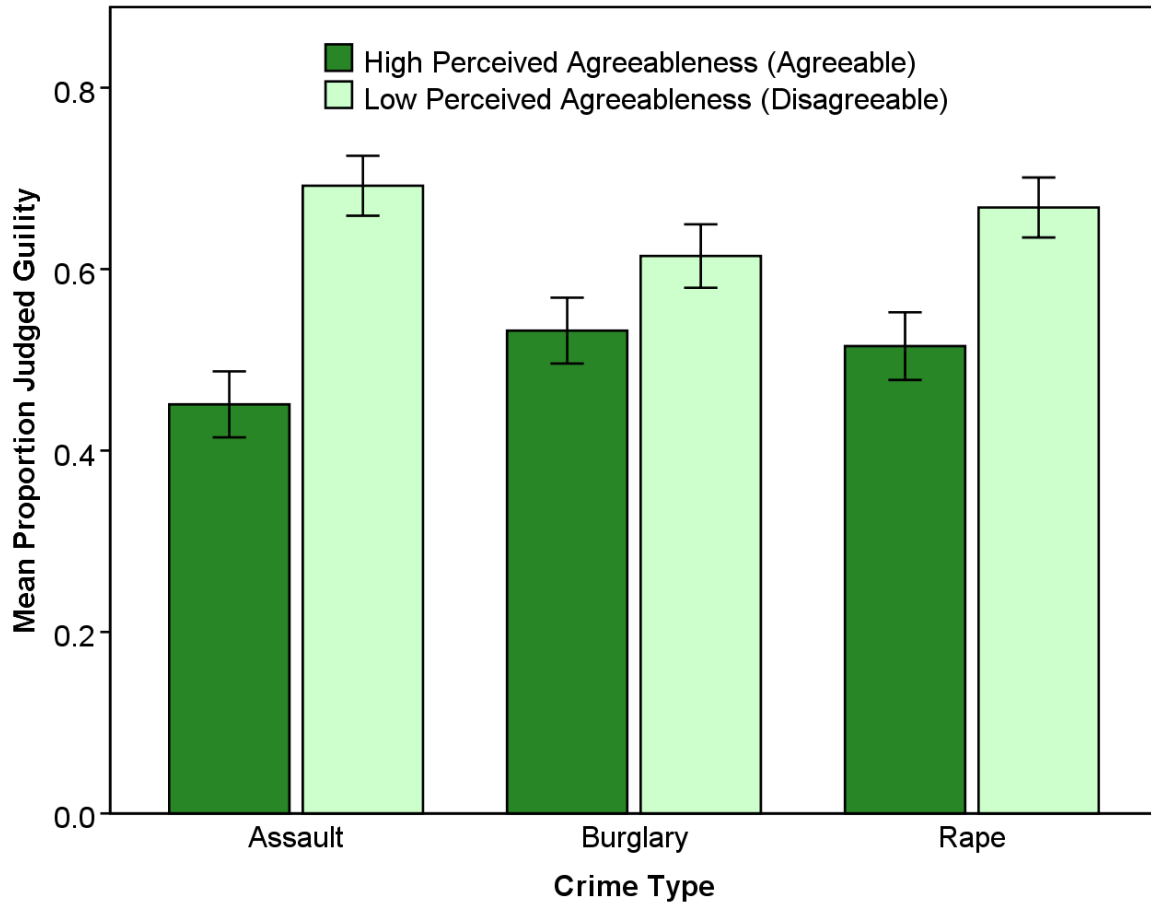

Figure 3: Effects perceived agreeableness manipulations on probability of being judged guilty of a crime, collapsed across morphological masculinity manipulations, (error bars = 95% CI). Aggregate data for 4428 judgements made by 369 participants viewing 12 photos each.

Confidence in judgements

In the preceding analyses, the 4428 trials completed by the sample of 369 participants judging 12 hypothetical crimes each, were treated as a pooled dataset. Due to the random presentation procedure, and small number of trials per participant (12), only 310 participants judged faces from all 4 face categories. Of these, 158 participants gave at least one guilty verdict to each of the 4 face types – i.e. a) feminized-agreeable; b) feminized-disagreeable; c) masculinized-agreeable; d) masculinized-disagreeable. To examine the effects of the appearance dimensions on participant confidence in the "guilty" judgements they gave, data from these 158 participants were examined. For each participant, a mean confidence rating was calculated for all "guilty" verdicts for each face type. Overall means for each category, shown in Figure 4, were a) feminized-agreeable ($M = 3.97$, $SD = 1.47$); b) feminized-disagreeable ($M = 4.24$, $SD = 1.50$); c) masculinized-agreeable ($M = 3.95$, $SD = 1.51$); d) masculinized-disagreeable ($M = 4.15$, $SD = 1.50$).

A 2 x 2 repeated measures ANOVA was conducted including these 158 participants, with facial morphological masculinity (masculine-feminine) and perceived agreeableness (agreeable-disagreeable) as within-subject factors, and the mean confidence rating given by a participant in a face category as the dependent. This revealed a significant main effect of perceived agreeableness, $F(1, 157) = 11.70$, $p < .001$, $\eta_p^2 = .069$, with low agreeableness being associated with greater confidence in guilt (shown in Figure 4). However, there was no significant main effect of facial morphological masculinity, $F(1, 157) = 0.83$, $p = .365$, $\eta_p^2 = .005$, and no significant agreeableness x masculinity interaction, $F(1, 157) = 0.21$, $p = .648$, $\eta_p^2 = .001$.
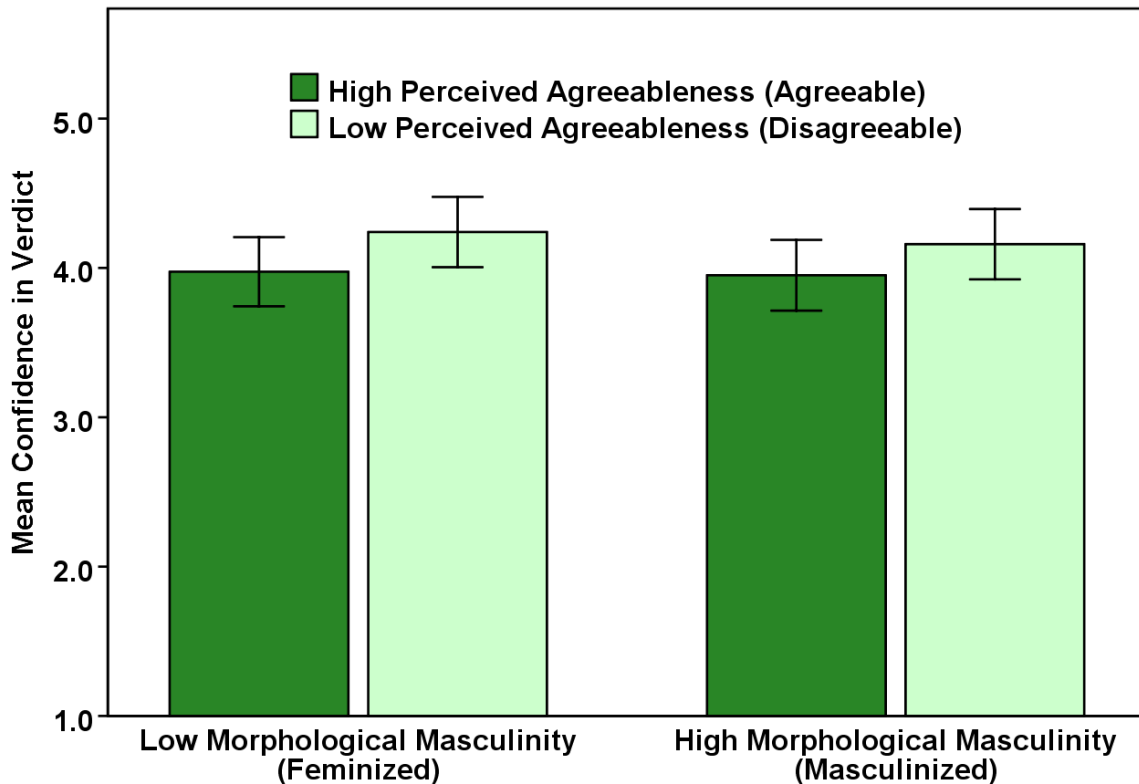
Figure 4: Effects of facial morphological masculinity and agreeableness manipulations on confidence in "guilty" judgements for 158 participants who gave at least one guilty verdict to each of the 4 face types (error bars = 95% CI).

A similar analysis was carried out for confidence ratings given for "not guilty" verdicts given to each face type. Overall means for each category, shown in Figure 5, were a) feminized-agreeable ($M = 3.64$, $SD = 1.63$); b) feminized-disagreeable ($M = 3.26$, $SD = 1.61$); c) masculinized-agreeable ($M = 3.63$, $SD = 1.65$); d) masculinized-disagreeable ($M = 3.44$, $SD = 1.71$). This could only include the 76 participants who gave at least one "not guilty" verdict to each of the 4 face types. This again revealed a significant main effect of perceived agreeableness, $F(1, 75) = 10.34$, $p = .002$, $\eta_p^2 = .121$, but this time with high agreeableness being associated with greater confidence in judgements that the target was "not guilty"

(shown in Figure 5). However, there was again no significant main effect of facial morphological masculinity, $F(1, 75) = 1.27$, $p = .263$, $\eta_p^2 = .017$, and no significant agreeableness x masculinity interaction, $F(1, 75) = 0.92$, $p = .341$, $\eta_p^2 < .012$.
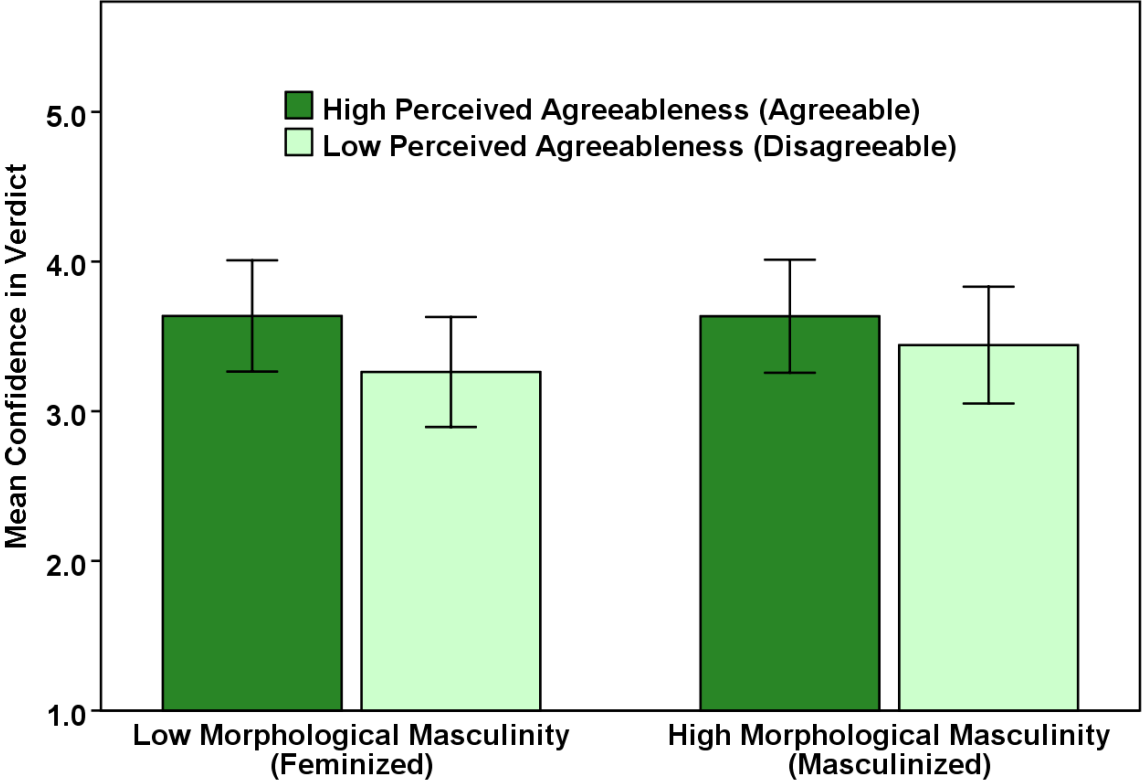

Figure 5: Effects of facial morphological masculinity and agreeableness manipulations on confidence in "not guilty" judgements for 76 participants who gave at least one not guilty verdict to each of the 4 face types (error bars = 95% CI).

Discussion

Facial appearance was shown to have a significant effect on the probability of being judged to be guilty of a crime and this was the case for all three crime types (assault, burglary and rape). For each crime type, manipulations of perceived agreeableness had large effects on the probability of being judged guilty – whereas manipulations of facial morphological masculinity did not. This was demonstrated by the simple chi-square analyses pooling data from all participants, which revealed a large effect for perceived agreeableness manipulations, but no significant effect for morphological masculinity manipulations. In addition, participants expressed greater confidence in their judgements of guilt when perceived agreeableness was low, and greater confidence in judgements of innocence when perceived agreeableness was high. In contrast, morphological masculinity manipulations had no effect on participant confidence in their judgements.

These findings are broadly consistent with previous research that has shown, using various methods, that facial appearance can have important effects on the probability of being judged guilty of a crime (e.g. Darby & Jeffers, 1988; Efran, 1974; for a meta-analysis see Mazzella & Feingold, 1994). In particular, the findings concur closely with those of Wilson & Rule (2015) who found that perceived trustworthiness in faces predicted severity of sentencing. There is strong conceptual overlap between the dimension of trustworthiness that they assessed in naturalistic stimuli, and the agreeableness dimension that were manipulated experimentally in the present study.

The lack of an effect for facial masculinity manipulations, however, is not consistent with the recent findings of Ward et al. (2012) and Estrada-Reynolds et al (2017). This discrepancy is likely due to some important methodological differences. Those previous studies both assessed facial masculinity through subjective ratings of the trait – i.e. they focused on perceived masculinity. In the case of Ward et al. (2012) participants judging

22

criminal guilt were shown faces that had been rated as exhibiting either low, moderate or high levels of masculinity. Estrada-Reynolds et al (2017) also used as stimuli, faces that were rated as high or low in masculinity and also used morphing techniques to create moderately masculine faces. As noted early, these methods have the potential to introduce some circularity to the evidence. If participants do associated traits such as dishonesty and disagreeableness with "masculinity" – then faces that are perceived as more dishonest and disagreeable may get rated as more masculine. In which case, it should not be surprising that such faces are then also perceived as more likely to commit crimes.

The finding that facial morphological masculinity does not appear to have a major effect on interpersonal perception when considered along with another more salient aspect of facial appearance, is not unprecedented. Research on determinants of male facial attractiveness has shown that, when stimuli are created that vary in morphological masculinity only, then this dimension can have important effects on an aspect of interpersonal perception - attractiveness judgements (e.g. DeBruine, Jones, Smith, & Little, 2010). However, morphological masculinity seems to be a less important determinant of attractiveness in faces that also vary on other dimensions such as skin colour (e.g. Scott et al., 2010; Stephen et al., 2012).

The failure to detect an effect of masculinity manipulations, when these occurred alongside manipulations of another trait, was probably not due to the size of the masculinity manipulation used. The magnitude of the manipulation employed (±50% masculinity transforms) has been commonly used in previous studies that have demonstrated important effects of masculinity on attractiveness (see DeBruine et al., 2010) and evaluations of personality (e.g. Kruger, 2006) and attractiveness (e.g. Welling et al, 2007). But in those examples, it was only masculinity that was manipulated – making it the most salient dimension of variation. In the present study, as in Scott et al. (2010) and Stephen et al. (2012)

the salience of masculinity variation appears to be reduced when other aspects of facial appearance are varied simultaneously – as would occur in natural settings. While the generalisability of the present study may be somewhat limited through the use of only White European face stimuli, it should be noted that a similar lack of salience of facial masculinity reported by Stephen et al. (2012) was found for both the Black African and White European face stimuli in that study. However, in a cross-cultural study of 12 populations, Scott et al (2014) found, for people judging attractiveness, important variation in preferences for morphological masculinity in faces (using stimuli produced using the technique we have used). Consequently, it would be interesting to examine whether there is also substantial cross-cultural variation in the extent to which masculinity is associated with judgements of criminal guilt.

Regarding the ecological validity of the procedure used in the present study, while vignette approaches have some limitations, they are commonly used in "mock jury" studies (Mazzella & Feingold, 1994) and have some important advantages. As noted by Bieneck (2015), they allow for large scale surveys (like the present study), and importantly are non-directive. That is, they make information available to participants, but it is the participants who determine which pieces of information are actually used in forming their judgements.

In order to focus on initial impression formation by individuals, the present study did not involve a full "mock jury" procedure that attempted to mimic all elements of criminal proceedings. Instead, through the use of photographs accompanied by short vignettes, it attempted to replicate the everyday situations where people may make rapid judgements about the likelihood of guilt after seeing reports about crimes in newspapers, on television, or on news websites. Nevertheless, the findings may have some implications for what does occur during courtroom proceedings – as they provide further clear evidence that aspects of facial appearance can have important effects on the likelihood of being perceived as being guilty of

committing a criminal offence. Of course, as in full "mock jury" studies, these were hypothetical judgements regarding hypothetical crimes. Nonetheless, the findings may have important implications as observational studies suggest that facial appearance can influence the outcome of real criminal cases (e.g. Stewart, 1980; 1985; Wilson & Rule, 2015).

      With evidence accumulating suggesting that judgements about people based on their facial appearance alone might influence judicial proceedings, in theory this could inspire modifications to court procedures to eliminate this source of bias (e.g. preventing a jury from seeing a defendant's face). However, the idea that judges and juries need to be able to examine the "demeanour" of witnesses (and defendants) is central to the adversarial legal process (Wellborn, 1991), and this view has informed recent decisions about whether witnesses can wear face coverings during court proceedings (Naudé, 2013). Nevertheless, even if courts insist that faces are available to be judged, knowledge that this can introduce bias to proceedings could still be utilized to make a trial fairer. For example, juries could be cautioned about the risks of bias, and the potential for judgements to be influenced by aspects of facial appearance that are not reliable cues to guilt or innocence.

References

Alharbi, S. A. H., Holzleitner, I. J., Lee, A. J., Saribay, S. A., & Jones, B. C. (2020). Women's Preferences for Sexual Dimorphism in Faces: Data from a Sample of Arab Women. *Evolutionary Psychological Science*. https://doi.org/10.1007/s40806-020-00244-y

Berry, D. S., & McArthur, L. Z. (1985). Some components and consequences of a babyface. *Journal of Personality and Social Psychology, 48*(2), 312-323. http://dx.doi.org/10.1037/0022-3514.48.2.312

Bieneck, S. (Ed.). (2009). How adequate is the vignette technique as a research tool for psycho-legal research? In M. E. Oswald, S. Bieneck, & J. Hupfeld-Heinemann (Eds.), *Social psychology of punishment of crime* (p. 255–271). John Wiley & Sons Ltd.

Brown, M., Sacco, D. F., Lolley, K. P., & Block, D. (2017). Facing the implications: Danger-ous world beliefs differentially predict men and women's aversion to facially communi-cated psychopathy. *Personality and Individual Differences, 116*, 1– 5. https://doi.org/10.1016/j.paid.2017.04.018

Brown, M., Sacco, D. F., & Medlin, M. M. (2019). Approaching extraverts: Socially excluded men prefer extraverted faces. *Personality and Individual Differences, 137*, 198– 203. https://doi.org/10.1016/j.paid.2018.09.007

Darby, B., & Jeffers, D. (1988). The effects of defendant and juror attractiveness on simulated courtroom trial decisions. *Social Behavior and Personality: An international journal*, *16*(1), 39-50. http://dx.doi.org/10.2224/sbp.1988.16.1.39

DeBruine, L. M., & Jones, B. C. (2017). Face Research Lab London Set (Version 3). figshare. http://dx.doi.org/10.6084/m9.figshare.5047666.v3

Debruine, L. M., Jones, B. C., Crawford, J. R., Welling, L. L. M., & Little, A. C. (2010). The health of a nation predicts their mate preferences: Cross-cultural variation in women's preferences for masculinized male faces. *Proceedings of the Royal Society B: Biological Sciences, 277*(1692), 2405-2410. http://dx.doi.org/10.1098/rspb.2009.2184

DeBruine, L. M., Jones, B. C., Smith, F. G., & Little, A. C. (2010). Are attractive men's faces masculine or feminine? The importance of controlling confounds in face stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 36*(3), 751-758. http://dx.doi.org/10.1037/a0016457

DeBruine, L. M., Jones, B. C., Little, A. C., Boothroyd, L. G., Perrett, D. I., Penton-Voak, I. S., et al. (2006). Correlated preferences for facial masculinity and ideal or actual partner's masculinity. *Proceedings of the Royal Society of London B, Biological Sciences, 273*(1592), 1355–1360. http://dx.doi.org/10.1098/rspb.2005.3445

DeSantis, A., & Kayson, W. A. (1997). Defendants' characteristics of attractiveness, race, and sex and sentencing decisions. *Psychological Reports, 81*(2), 679-683. http://dx.doi.org/10.2466/pr0.1997.81.2.679

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41*, 417-440. http://dx.doi.org/10.1146/annurev.ps.41.020190.002221

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology, 24*(3), 285-290. http://dx.doi.org/10.1037/h0033731

Efran, M. G. (1974). The effect of physical appearance on the judgment of guilt, interpersonal attraction, and severity of recommended punishment in a simulated jury task. *Journal of Research in Personality, 8*(1), 45-54. http://dx.doi.org/10.1016/0092-6566(74)90044-0

Estrada-Reynolds, V. C., Reynolds, J. J., McCrea, S. M., & Freng, S. (2017). I don't like the cut of your jib: Perceived facial masculinity as a cue to criminality. *Psychiatry, Psychology and Law, 24*(3), 392-409. http://dx.doi.org/10.1080/13218719.2016.1247639

Flowe, H. D. (2012). Do characteristics of faces that convey trustworthiness and dominance underlie perceptions of criminality? *PLoS ONE, 7*(6) http://dx.doi.org/10.1371/journal.pone.0037253

Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology, 78*(5), 837-852. http://dx.doi.org/10.1037/0022-3514.78.5.837

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). Madison, CT, US: Psychosocial Press.

Huelsnitz, C. O., Neel, R., & Human, L. J. (2020). Accuracy in Perceptions of Fundamental Social Motives: Comparisons to Perceptions of Big Five Traits and Associations With Friendship Quality. *Personality and Social Psychology Bulletin, 46*(1), 3–19. https://doi.org/10.1177/0146167219838546

John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), Handbook of personality: Theory and research (Vol. 2, pp. 102–138). New York: Guilford Press

Johnston, V. S., Hagel, R., Franklin, M., Fink, B., & Grammer, K. (2001). Male facial attractiveness: Evidence for hormone-mediated adaptive design. *Evolution and Human Behavior, 22*(4), 251-267. http://dx.doi.org/10.1016/S1090-5138(01)00066-6

Kruger, D. J. (2006). Male facial masculinity influences attributions of personality and

reproductive strategy. *Personal Relationships, 13*(4), 451-463.

http://dx.doi.org/10.1111/j.1475-6811.2006.00129.x

Kulka, R. A., & Kessler, J. B. (1978). Is justice really blind?–The influence of litigant

physical attractiveness on juridical judgment. *Journal of Applied Social Psychology, 8*(4),

366-381. http://dx.doi.org/10.1111/j.1559-1816.1978.tb00790.x

Leventhal, G., & Krate, R. (1977). Physical attractiveness and severity of sentencing.

Psychological Reports, 40(1), 315-318. http://dx.doi.org/10.2466/pr0.1977.40.1.315

Liggett, J. (1974). *The human face*. New York: Stein & Day.

Marcinkowski, F., Lünich, M., & Starke, C. (2018). Spontaneous trait inferences from

candidates' faces: The impact of the face effect on election outcomes in Germany. *Acta

Politca, 53*(2), 231-247. http://dx.doi.org/10.1057/s41269-017-0048-y

Maurer-Fazio, M., & Lei, L. (2015). "As rare as a panda": How facial attractiveness, gender,

and occupation affect interview callbacks at Chinese firms. *International Journal of

Manpower, 36*(1), 68-85. http://dx.doi.org/10.1108/IJM-12-2014-0258

Mayer, J. D., Phillips, K. G., & Barry, A. (2015). Getting the Message: The Adaptive Poten-

tial of Interpersonal Judgments. Review of General Psychology, 19(1), 39–51.

https://doi.org/10.1037/gpr0000025

Mazzella, R., & Feingold, A. (1994). The effects of physical attractiveness, race,

socioeconomic status, and gender of defendants and victims on judgments of mock

jurors: A Meta-Analysis. *Journal of Applied Social Psychology, 24*(15), 1315-1338.

http://dx.doi.org/10.1111/j.1559-1816.1994.tb01552.x

McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception.

Psychological Review, 90(3), 215–238. https://doi.org/10.1037/0033-295X.90.3.215

Medlin, M.M., Sacco, D.F. & Brown, M. (2020). The Relation Between Narcissistic

Personality Traits and Accurate Identification of, and Preference for, Facially

Communicated Narcissism. *Evolutionary Psychological Science* **6,** 166–173.

https://doi.org/10.1007/s40806-019-00224-x

Mitteroecker, P., Windhager, S., Müller, G. B., & Schaefer, K. (2015). The morphometrics of

"masculinity" in human faces. *PLoS ONE, 10*(2), e0118374.

http://dx.doi.org/10.1371/journal.pone.0118374

Naudé, B. (2013). Face-coverings, demeanour evidence and the right to a fair trial: lessons

from the USA and Canada. The Comparative and International Law Journal of Southern

Africa, 46(2), 166-183.

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. http://dx.doi.org/10.1073/pnas.0805664105

Penton-Voak, I. S., Pound, N., Little, A. C., & Perrett, D. I. (2006). Personality judgments

from natural and composite facial images: More evidence for a "kernel of truth" in social

perception. *Social Cognition, 24*(5), 607-640.

http://dx.doi.org/10.1521/soco.2006.24.5.607

Perrett, D. I., Lee, K. J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D. M., Henzi., S.

P., Castles, D. L. & Akamatsu, S. (1998). Effects of sexual dimorphism on facial

attractiveness. *Nature, 394*(6696), 884-887. http://dx.doi.org/10.1038/29772

Rowe, D. C., Flannery, A. T., & Flannery, D. J. (1995). Sex differences in crime: Do means

and within-sex variation have similar causes? *Journal of Research in Crime and

Delinquency, 32*(1), 84-100. http://dx.doi.org/10.1177/0022427895032001004

Rowland, D. A., & Perrett, D. I. (1995). Manipulating facial appearance through shape and

    color. *IEEE Computer Graphics and Applications, 15*, 70–76.

    http://dx.doi.org/10.1109/38.403830

Sacco, D. F., Brown, M. (2018). The face of personality: Adaptive inferences from facial cues

    are moderated by perceiver personality and motives. *Soc Personal Psychol Compass,*

    12(8):e12410. https://doi.org/10.1111/spc3.12410

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://dx.doi.org/10.1371/journal.pone.0112157

Scott, I. M., Clark, A. P., Josephson, S. C., Boyette, A. H., Cuthill, I. C., Fried, R. L., . . .

    Penton-Voak, I. S. (2014). Human preferences for sexually dimorphic faces may be

    evolutionarily novel. *Proceedings of the National Academy of Sciences of the United*

    States of America, 111(40), 14388-14393. http://dx.doi.org/10.1073/pnas.1409643111

Scott, I. M. L., Pound, N., Stephen, I. D., Clark, A. P., & Penton-Voak, I. S. (2010). Does

    masculinity matter? The contribution of masculine face shape to male attractiveness in

    humans. *PLoS ONE, 5*(10), e13585. http://dx.doi.org/10.1371/journal.pone.0013585

Stephan, C. N., Penton-Voak, I. S., Perrett, D. I., Tiddeman, B. P., Clement, J. G., &

    Henneberg, M. (2005). Two-dimensional computer generated average human face

    morphology and facial approximation. In: J. G. Clement & M. K. Marks (Eds.),

    Computer Graphic Facial Reconstruction (pp. 105–127). Burlington, VT: Elsevier

    Academic Press.

Stephen, I. D., Scott, I. M. L., Coetzee, V., Pound, N., Perrett, D. I., & Penton-Voak, I. S.

    (2012). Cross-cultural effects of color, but not morphological masculinity, on perceived

attractiveness of men's faces. *Evolution and Human Behavior, 33*(4), 260-267.

http://dx.doi.org/10.1016/j.evolhumbehav.2011.10.003

Stewart, J. E. (1980). Defendant's Attractiveness as a Factor in the Outcome of Criminal

Trials: An Observational Study. *Journal of Applied Social Psychology, 10*(4), 348-361.

http://dx.doi.org/10.1111/j.1559-1816.1980.tb00715.x

Stewart, J. E. (1985). Appearance and Punishment: The Attraction-Leniency Effect in the

Courtroom. *The Journal of Social Psychology, 125*(3), 373-378,

http://dx.doi.org/10.1080/00224545.1985.9922900

Stirrat, M., & Perrett, D. I. (2010). Valid Facial Cues to Cooperation and Trust: Male Facial

Width and Trustworthiness. *Psychological Science, 21*(3), 349–354.

https://doi.org/10.1177/0956797610362647

Sutherland, C. A. M., Rhodes, G., & Young, A. W. (2017). Facial image manipulation: A tool

for investigating social perception. *Social Psychological and Personality Science, 8*(5),

538–551. http://dx.doi.org/10.1177/1948550617697176

Tiddeman, B., Burt, D. M., and Perrett, D. I. (2001). Prototyping and transforming facial

textures for perception research. *IEEE Comput. Graph. Appl. 21*, 42–50.

http://dx.doi.org/10.1109/38.946630

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence

from faces predict election outcomes. *Science, 308*(5728), 1623-1626.

http://dx.doi.org/10.1126/science.1110589

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation

of faces on social dimensions. *Trends in Cognitive Sciences, 12*(12), 455-460.

http://dx.doi.org/10.1016/j.tics.2008.10.001

Todorov, A. T., Said, C. C., & Verosky, S. C. (2012). Personality impressions from facial appearance. In A. Calder, G. Rhodes, M. Johnson, & J. V. Haxby (Eds.), *Oxford Handbook of Face Perception.* (pp. 630 – 651). Oxford: Oxford University Press.

Ward, C., Flowe, H., & Humphries, J. (2012). The effects of masculinity and suspect gender on perceptions of guilt. *Applied Cognitive Psychology, 26*(3), 482-488. http://dx.doi.org/10.1002/acp.2823

Wellborn III, O. G. (1991). Demeanor. *Cornell Law Review, 76*(5), 1075-1105.

Welling, L. L. M., Jones, B. C., DeBruine, L. M., Conway, C. A., Law Smith, M. J., Little, A. C., Feinberg, D.R., Sharp, M.A., & Al-Dujaili, E. A. S. (2007). Raised salivary testosterone in women is associated with increased attraction to masculine faces. *Hormones and Behavior, 52*(2), 156-161. http://dx.doi.org/10.1016/j.yhbeh.2007.01.010

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*(7), 592-598. http://dx.doi.org/10.1111/j.1467-9280.2006.01750.x

Wilson, J. P., & Rule, N. O. (2015). Facial Trustworthiness Predicts Extreme Criminal-Sentencing Outcomes. *Psychological Science, 26*(8), 1325–1331. https://doi.org/10.1177/0956797615590992

Wilson, J. Q., & Herrnstein, R. J. (1985). *Crime and human nature*. New York, NY, US: Simon & Schuster.

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. http://dx.doi.org/10.1207/s15327957pspr0103_2

Zebrowitz, L. A., & Franklin Jr., R. G. (2014). The attractiveness halo effect and the babyface stereotype in older and younger adults: Similarities, own-age accentuation, and older

adult positivity effects. *Experimental Aging Research, 40*(3), 375-393.

http://dx.doi.org/10.1080/0361073X.2014.897151

Zebrowitz, L. A., & McDonald, S. M. (1991). The impact of litigants' baby-facedness and attractiveness on adjudications in small claims courts. *Law and Human Behavior, 15*(6), 603-623. http://dx.doi.org/10.1007/BF01065855

Zebrowitz, L. A., & Montepare, J. M. (1992). Impressions of babyfaced individuals across the life span. *Developmental Psychology, 28*(6), 1143-1152. http://dx.doi.org/10.1037/0012-1649.28.6.1143

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. http://dx.doi.org/10.1177/01461672962212006

Zebrowitz, L. A., Wang, R., Bronstad, P. M., Eisenberg, D., Undurraga, E., Reyes-García, V., & Godoy, R. (2012). First impressions from faces among U.S. and culturally isolated Tsimane' people in the Bolivian rainforest. *Journal of Cross-Cultural Psychology, 4*3(1):119–134. http://dx.doi.org/10.1177/0022022111411386

Tests for effects of face manipulations on proportion of guilty verdicts for data split by gender of participants making the judgements.

The overall proportion of guilty judgements given was similar for male (42.1%) and female (41.7%) participants. Supplementary Table 1 shows the percentage of each face type judged guilty for each crime type split by gender of rater. Collapsing across agreeableness manipulations, a series of 2 x 2 (Masculinity x Verdict) chi-square tests (one for each crime type) showed that facial masculinity had no significant effects on perceptions of guilt for any crime type for the data from male participants (all $\chi^2 < 0.83$; $df = 1$; $p > .36$) and female participants (all $\chi^2 < 0.42$; $df = 1$; $p > .52$). In contrast, however, collapsing across masculinity manipulations, a series of 2 x 2 (Agreeableness x Verdict) chi-square tests (one for each crime type) showed for data from male participants, facial agreeableness had significant effects on perceptions of guilt for assault ($\chi^2 = 33.00$; $df = 1$; $p < .001$), burglary ($\chi^2 = 4.79$; $df = 1$; $p = .029$) and rape ($\chi^2 = 8.71$; $df = 1$; $p = .003$) from male participants. A similar pattern was seen for data from female participants for assault ($\chi^2 = 53.05$; $df = 1$; $p < .001$), burglary ($\chi^2 = 4.73$; $df = 1$; $p = .030$) and rape ($\chi^2 = 27.83$; $df = 1$; $p < .001$).

Supplementary Table 1: Percentage of each face type judged guilty for each crime type split by gender of rater.

	Low Morphological Masculinity (Feminized)		High Morphological Masculinity (Masculinized)	
	a) Agreeable	b) Disagreeable	c) Agreeable	d) Disagreeable
Male raters				
Assault	42.6%	67.2%	48.6%	67.3%
Burglary	53.1%	65.8%	53.3%	57.8%
Rape	53.9%	65.5%	52.5%	63.4%
Female raters				
Assault	42.7%	72.9%	47.7%	67.9%
Burglary	55.7%	57.1%	51.6%	65.0%
Rape	51.4%	68.4%	49.3%	68.7%