Is sexual dimorphism under present selection in humans?

LINDA HELENA LIDBORG

LIDBORG, LINDA HELENA (2022) Is sexual dimorphism under present selection in humans? Doctoral thesis, Durham University.

Is sexual dimorphism under present selection in humans?

A thesis submitted for the degree of Doctor of Philosophy

in the Department of Psychology

Durham University

Linda Helena Lidborg

2022

Abstract

Like most mammalian species, humans show sexual dimorphism: the two sexes exhibit sex-typical morphological traits. Humans show dimorphism in the facial structure, body size/shape, physical strength, the amount of and distribution of muscle and fat mass, voice pitch, finger ratios, and the growth of facial and body hair. Such dimorphisms are argued to be sexually selected. This thesis focussed on testing biological fitness benefits as a function of exaggerated sexual dimorphism (masculinity in men; femininity in women) in humans. In Chapter 2, a comprehensive meta-analysis showed that men's body masculinity (increased strength/muscularity) significantly predicted greater mating/reproductive success. A more masculine voice pitch, increased height, and higher testosterone levels positively predicted mating, but not reproduction. Facial masculinity and finger ratios did not significantly predict either. Chapter 3 focussed on women's traits. Novel analyses of two archival datasets indicated that women with more appealing faces reported fewer births, but greater offspring survival. Furthermore, a systematic literature review showed no consistent associations between women's traits and reproductive outcomes, and meta-analyses of men's and women's fitness as a function of facial attractiveness showed no relationships. In Chapter 4, we tested whether more dimorphic parents had offspring who showed better health and/or earlier sexual onset in two samples. We detected very few significant associations, whereby fathers with greater facial dominance, strength, and height had offspring in better health, and more facially attractive and healthy-looking fathers had offspring who showed earlier sexual onset. Meta-analysing the relationships between parental facial traits and offspring health across the two samples revealed effects close to zero. Lastly, in Chapter 5 we tested whether more masculine men had partners who were more feminine/attractive, healthier, and/or had higher mate value in two U.K. samples. Men with more masculine bodies had more facially appealing partners, and more facially appealing men had higher partner-rated mate value but self-reported somewhat worse health. Except for the meta-analytic effects in Chapter 2, none of the significant effects here remained after correcting for multiple comparisons. The only robust associations we detected were thus with respect to masculinity in men’s bodies increasing fitness outcomes. Overall, the results from this thesis give very limited support to the notion that other dimorphic traits confer men and women fitness benefits. The implications for human sexual selection models are discussed.

Table of contents

Chapter 5. Male morphology and attractiveness as indicators of partner quality …….…..132

Chapter 6. General discussion ………..…………………………………………………158

List of tables

List of figures

Foreword

Throughout the analyses in this thesis, outliers were retained. Corrections for multiple corrections were done throughout by computing *q*-values (associations where *q*-values were below .05 were considered significant). *Q*-value computations were done using all *p* values obtained from all analyses carried out within each chapter. We used the *qvalue* R package for *q*-value computation (Storey et al., 2017).

Submitted work

The entirety of Chapter 2 has been submitted for publication and is presented as submitted.

Declaration

I declare that no part of the material presented in this thesis has previously been submitted by me for a degree in this or in any other university. The archival dataset from the Agta, reported in Chapter 3, is open data. The archival data from the Turkish sample, reported in Chapters 3 and 4, is data for which the supervisor is the data controller. The data from the Tees Valley Baby sample, reported in Chapters 3 and 4, is joint data partially collected by me. In all other cases, the work of others has been acknowledged and referenced appropriately.

Statement of copyright

Acknowledgements

Firstly, the biggest of thanks to my supervisor, Lynda Boothroyd. I cannot thank you enough for your support, help, advice, and kindness (and for being super speedy with feedback, not to mention always being available for a message on Twitter!). My primary memory of my PhD journey will always be the many an hour spent in your office, discussing *the* meta-analysis. A fond memory, in retrospect (but I’m also quite relieved it’s over..!). Thank you; you truly put the ‘super’ in ‘supervisor’!

Thank you to Mike Burt, for your help setting up the lab-based data collections.

An enormous thank you to my mum, for always being there for me and supporting me (not to mention, looking after my dogs for me when I’ve been off on my academic adventures).

Thank you to my dad. I wish you could have been here for this.

Thank you to the best PhD buddies/support team, Becca, Emine, Liz, and Naz; thank you to my desk neighbour and plant caretaker, Diane; and thank you to Anders, Fabienne, and Jonathan, for helpful comments on my chapters.

Thank you to my very dearest friend, Kristina, for being an all-over great friend, PhD buddy, and fellow exchanger of cute dog pictures.

And, lastly, thank you to the very best of PhD puppies, Ziggy and Poppy, for putting a smile on my face every day and for keeping me company during many, long hours of writing this thesis.

Chapter 1.

General introduction

1.1. Sexual dimorphism

Across both animal and plant species, males and females within the same species may differ from one another to varying degrees. Such sex differences can refer to so-called ‘primary’ differences − differences in traits related to mating and reproduction (which may be inevitable in a sexually reproducing species) – and ‘secondary’ differences, which are unrelated to mating/reproduction (Plavcan, 2001). When the males and females of a species exhibit secondary sex differences, this is referred to as ‘sexual dimorphism’. Sexual dimorphism can manifest in a range of phenotypic traits, including both morphologies and behaviours (Barnard, 2004). Morphological dimorphism in animals includes – but is not limited to − differences in body size, fur growth, colours and markings, and weaponry such as antlers and enlarged canines. Behavioural dimorphism might include engagement in courtship rituals, physical contests, and threat displays. Crucially, dimorphic traits are either only present or are exaggerated in one sex, and not the other. Traits that are exaggerated in males are typically referred to a ‘masculine’, whereas traits typical for females are referred to as ‘feminine’. Sexually dimorphic traits usually develop at sexual maturity. In contrast, sexual monomorphism refers to when the two sexes show highly similar or virtually identical phenotypes.

How do sexually dimorphic traits evolve? Ultimately, anisogamy (i.e., asymmetry in the size of sex cell production) facilitates the evolution of sexual dimorphism (Smith & Maynard-Smith, 1978), since it necessitates greater maternal investment in offspring while simultaneously making it possible for males to desert following copulation (particularly in species where gestation is internal). This asymmetry in parental investment tends to result in a male-biased operational sex ratio, whereby adult males are always sexually receptive whereas some females are not due to already being pregnant or by currently investing in offspring (Mitani et al., 1996). This, in turn, results in females being a limiting factor for male reproduction, and males being subject to relatively greater competition for mates to fertilise (Emlen & Oring, 1977). In contrast, the primary constraint on female reproduction is not access to mates, but the rate at which females can produce (surviving) offspring (Andersson, 1994). Thus, anisogamy underlies sexual selection: within-species selection for traits that facilitate increased reproductive output.

The evolution of specific sexually dimorphic traits is usually attributed to one of the two main components of sexual selection: inter- or intrasexual selection (Barnard, 2004). In intersexual selection, one sex − typically females − chooses whom to mate with and preferentially mates with individuals who exhibit certain traits (i.e., so-called *female choice*). In contrast, intrasexual selection refers to within-sex competition – typically between males – either for direct sexual access to mates or for territory/dominance positions, which come with mating rights (i.e., so-called *male-male competition* or *contest competition*). Female choice is prevalent in, for instance, many bird species, where males have evolved conspicuous ornaments and courtship displays in order to attract females (Zahavi, 1975). An example of this is seen in peafowl, where the peacock has evolved elaborate, vividly coloured tail plumage; the peahen, in comparison, remains dull in appearance. Male-male competition, in contrast, tends to result in the evolution of size and strength dimorphism as well as weaponry (reviewed in Isaac, 2005). Unsurprisingly, monomorphism is usually seen in monogamous species, and dimorphism is typically most pronounced in species with strong mate competition such as in polygynous mating systems (Clutton-Brock & Harvey, 1977).

Sexual dimorphism can thus be found in a range of species and take different forms, depending (in part) on the mating system of the species. Extensive mammalian male-male competition has resulted in pronounced sexual dimorphism in most mammalian species (Lindenfors et al., 2007), often taking the form of males being larger and stronger than females. Focussing on our primate relatives, sexual dimorphism is widespread (at least in anthropoids; monomorphism is more common in strepsirrhines and tarsiers: Plavcan, 2001) and varies from slight to extreme in magnitude. Some of the most striking examples of primate dimorphism can be seen in geladas, baboons, and gorillas. Most typically, sex differences in primates are seen with respect to body mass and canine size (although pelage and skin dimorphism occurs too: Plavcan, 2001). While male primates may show between 30-80% greater body mass than females, canine dimorphism can be even more extreme, ranging up to 400% greater in males in some primate species (Plavcan, 2001). Such dimorphisms are generally attributed to male-male competition (although it should be noted that the factors influencing the evolution of sexual dimorphism within a species are not limited to the mating system, and are rather varied and complex: Plavcan, 2011).

1.2. Sexual dimorphism in humans

Turning now from dimorphism in non-human animals to humans, our species exhibits sexual dimorphism in facial morphology (Samal et al., 2007), body size (Plavcan, 2001) and shape (Hughes & Gallup Jr, 2003; Singh, 1994), muscle mass (Lassek & Gaulin, 2009) and fat mass/distribution (Singh, 1994), physical strength (Lassek & Gaulin, 2009), facial and body hair (Dixson et al., 2005), voice pitch (Puts et al., 2012), and 2nd to 4th digit (2D:4D) ratios (Manning, 2002). Relative to non-human primates, humans show moderate levels of dimorphism, particularly with regards to body size (Plavcan, 2012). However, in non-human mammals, size dimorphism is typically strongly dependent upon overall body mass/weight. Humans, in contrast, show reduced mass/weight dimorphism due to human females carrying substantial amounts of body fat (Plavcan, 2012; Wells, 2007). Focussing exclusively on dimorphism in muscle/lean mass paints a different picture: here, humans show dimorphism comparable to other primates in magnitude (Plavcan, 2012).

This thesis will focus on dimorphic traits that are exaggerated in human males, i.e., masculinity. The development of masculine traits in men is generally influenced by prenatal and adolescent exposure to androgens, particularly testosterone. In the next section, the expression and development of masculine traits will be outlined.

1.3. Masculine traits in human males

1.3.1. Facial masculinity

Compared to women, men have more a more ‘robust’ facial morphology, with a pronounced brow ridge, a longer lower face, and wide mandibles, cheekbones, and chins (Swaddle & Reierson, 2002). In masculine faces, the eyes also appear smaller due to being more deeply set, the eyebrows are thicker, the lips thinner, and the nose larger (Kleisner et al., 2021; Rhodes, 2006). The more prominent lower part of the face may be further amplified by the testosterone-dependent growth of facial hair which is present in men from puberty (Dixson et al., 2005; Farthing et al., 1982). The more exaggerated these features are, the more masculine the face appears.

Facial dimorphism in human faces is thus present both in the skeletal structure and in soft tissues (Marečková et al., 2011). Prior to puberty, the facial morphology of boys and girls is similar but not identical (Bulygina et al., 2006; Snodell et al., 1993). Male and female faces start to diverge more substantially during puberty (Samal et al., 2007; Weston et al., 2007), and adult levels of facial dimorphism are reached towards the late teens (Snodell et al., 1993). The pubertal divergence of facial morphology is largely attributable to the surge in testosterone seen in boys at sexual maturity, stimulating skeletal growth (Verdonck et al., 1999; Whitehouse et al., 2015), although prenatal hormone levels may also influence later trait expression (Whitehouse et al., 2015). While it is thus established that developmental testosterone levels influence facial morphology, it has sometimes been claimed that masculinity in adult men's faces also correlates with their levels of circulating testosterone levels (Penton-Voak & Chen, 2004; Pound et al., 2009); adult masculine trait expression is thus often argued to function as a proxy for current testosterone levels. It should be noted here, however, that the relationship between facial masculinity and testosterone levels in adulthood is equivocal (Kordsmeyer et al., 2019; Peters et al., 2008).

Different methods are employed to measure facial masculinity (reviewed in Sanchez-Pages et al., 2014). The most prevalent method is using subjective third-party ratings, where observers are asked to rate how masculine a face looks, most commonly based on a static 2D facial photograph (e.g., Peters et al., 2008). More objective facial masculinity scores can also be acquired through the use of geometric morphometric software, which is used to assess facial sexual dimorphism according to predetermined landmarks. Rated and morphometric facial masculinity correlate, but not perfectly so (Boothroyd et al., 2013; Mitteroecker et al., 2015). Facial masculinity can also be indexed by specific craniofacial measures such as face width/lower face height, cheekbone prominence, and larger lower face/full face height (Lefevre et al., 2012), or the commonly used facial width-to-height ratio (fWHR: Geniole et al., 2015). Caveats with regards to fWHR is that it is not always found to be sexually dimorphic (Kordsmeyer et al., 2019; Lefevre et al., 2012), it may not exhibit the pubertal growth spurt which is characteristic of sexually dimorphic traits (Hodges-Simeon et al., 2016), and it is not consistently found to correlate with testosterone levels either in adolescence (Hodges-Simeon et al., 2016) or in adulthood (Kordsmeyer et al., 2019).

1.3.2. Body masculinity (strength, muscle mass, and body shape)

Some of the greatest dimorphisms in humans is seen in terms of strength and muscle mass. In a large sample of U.S. men and women, Lassek and Gaulin (2009) reported that, compared to women, men exhibited 61% more overall muscle mass, 78% more muscle mass in the upper arms, and 90% greater upper body strength, rendering the average man stronger than 99.9% of women. Strength in humans is often indexed by handgrip strength, which is typically measured by squeezing a handheld dynamometer. Handgrip strength correlates with overall muscle mass (Kallman et al., 1990) and total muscle strength (Wind et al., 2010), but is largely genetically driven (Reed et al., 1991). The bones of men are also larger and heavier, and their muscle attachment areas greater (Plavcan, 2012).

Dimorphism in handgrip strength is present, albeit small in magnitude, in childhood (Montalcini et al., 2016) but increases rapidly from puberty (Butterfield et al., 2009; Hodges-Simeon et al., 2016). Correlational evidence has linked muscularity and strength (Gettler et al., 2010) and fat-free mass (Lukas et al., 2004) to circulating testosterone levels in adulthood. Experimental studies show that clinically administered testosterone increases men’s muscle volume, strength, and fat-free mass; these changes correlate with gains in total and free testosterone concentrations (Bhasin et al., 2001). Such changes are augmented if combined with exercise, but occur even independently of it (Bhasin et al., 1996). It appears that testosterone increases muscle mass in healthy men by increasing muscle protein synthesis (Griggs et al., 1989); muscle protein synthesis drives adaptive responses to exercise, resulting in improved performance and increased muscle mass (Atherton & Smith, 2012). While such responses are of course influenced by the type of exercise undertaken, there are also considerable individual differences (partly genetically determined) in ‘responsiveness’ to exercise (Timmons, 2011). Thus, skeletal muscle volume, fat-free mass, and physical strength are not just dependent on testosterone for trait development in puberty, but are subject to continuous influence of testosterone over a man's lifetime.

Reflecting the sex differences in amount and distribution of muscle mass, men's bodies also differ in shape from women's. Men's bodies tend to have a V- or wedge-shape; their shoulders are broader in relation to their hips compared to women, showing a greater shoulder-to-hip ratio (Hughes & Gallup Jr, 2003; Singh, 1994). The shoulder-to-hip ratio also develops during puberty and has been linked to testosterone levels (Kasperk et al., 1997). A closely related measure, also contributing to the impression of a V-shaped torso, is a greater waist-to-chest ratio (Tovée et al., 1999; Weeden & Sabini, 2007).

1.3.3. 2D:4D ratios

The fingers on each hand are numbered from 1 (the thumb) to 5 (the little finger); digit ratios thus refer to the length of one finger relative to the length of another on the same hand (Lu et al., 2017). The ratio of the second to the fourth digit, the 2D:4D ratio, is sexually dimorphic, with men showing a lower ratio than women (Manning, 2002), particularly in the right hand (Hönekopp et al., 2006). It should be noted, however, that 2D:4D dimorphism may not be robust across cultures (Apicella et al., 2016).

Unlike other dimorphic traits, 2D:4D dimorphism appears to be influenced primarily by pre- and postnatal − rather than pubertal − testosterone exposure. It is present already at birth (Galis et al., 2010) and is largely stabilised within the first two years of life (Ventura et al., 2012). In mice, experimental manipulation of androgens versus oestrogens has been shown to increase and decrease growth of the fourth digit, respectively, resulting in masculinised versus feminised digit ratios (Zheng & Cohn, 2011). Elsewhere, however, the association between prenatal hormone exposure and 2D:4D dimorphism has been questioned (Hollier et al., 2015; Richards et al., 2019). Meta-analytic evidence suggests that 2D:4D ratios are not related to adult hormone levels (Hönekopp et al., 2007) and they are therefore – unlike other dimorphic traits − typically treated as proxies for prenatal rather than adult hormone profiles.

1.3.4. Voice pitch

The human voice comprises two acoustic components: formant (resonant) frequency and fundamental frequency, f0 (Titze, 2000). Fundamental frequency is what is commonly referred to as voice pitch (Atkinson et al., 2012), produced by vibration of the vocal folds and influenced by the vocal tract's size and shape (Evans et al., 2008). Typically, the acoustic properties of recorded voices are measured using specialised software (e.g., Suire et al., 2018); however, some researchers also measure the perceived correlate of voice pitch by using third-party ratings of the dimorphism of recorded voice clips (e.g., Hill et al., 2013).

In childhood, there is little difference between male and female voices. Under the influence of testosterone exposure in puberty the vocal fold length in boys increases, thus deepening the voice (Harries et al., 1998; Hodges-Simeon et al., 2016). In adulthood, men's voice pitch is close to six standard deviations lower than women's (Puts et al., 2012); this makes voice pitch one of the most dimorphic traits in humans. There is evidence suggesting that a low voice pitch may be related to high levels of adult salivary testosterone (Aung & Puts, 2020; Dabbs Jr & Mallinger, 1999; Evans et al., 2008; Puts et al., 2011; but see also Arnocky et al., 2018).

1.3.5. Height

In all studied human populations, men are, on average, taller than women. Adult sexual dimorphism in height equals around 7-8% (Gray & Wolfe, 1980). In childhood, girls are taller than boys, but this relationship is reversed once puberty is reached (Zheng et al., 2013).

Height development is strongly influenced by genetics (Allen et al., 2010), but it is also influenced by environmental factors such as nutrition (dos Santos et al., 2014).

In terms of the association between height and exposure to sex hormones, Hodges-Simeon and colleagues (2016) observed a positive relationship between height and testosterone in adolescent and young adult Tsimané males. The authors also reported that height showed a pubertal growth spurt similar to voice pitch and strength. Experimental studies investigating the relationship between endogenous androgen administration and height has shown conflicting findings. For example, Arslanian and Suprasongsin (1997) reported a significant increase in height following testosterone administration in boys with delayed puberty, whereas Blethen and colleagues (1984) observed that testosterone administration had no influence on height in adolescent boys.

1.3.6. Masculine trait intercorrelations and relationships with testosterone

To summarise, most masculine traits in men thus develop or become exaggerated following pubertal exposure to testosterone. While the described traits are sometimes treated as proxies for men’s adult testosterone levels, the evidence that adult trait expression indexes circulating testosterone levels in adulthood is mixed. It should also be noted that the association between testosterone levels in adolescence, when masculine traits generally develop, and in adulthood is extremely weak (van Bokhoven et al., 2006), calling into question the validity of regarding masculine traits as proxies of current hormone profiles.

Since all these traits are influenced by testosterone, it may be expected that they should be correlated. Again, the evidence for this is mixed. There is evidence suggesting that men’s facial masculinity may be related to indices of their body masculinity (e.g., Fink et al., 2007; Peters et al., 2008; van Dongen & Sprengers, 2012; Windhager et al., 2011), but see also null results reported by Hill and colleagues (2013) and van Dongen, (2014). As mentioned previously, Hodges-Simeon and colleagues (2016) found that height, voice pitch, and handgrip strength (but not fWHR) showed similar developmental trajectories as well as similar moderate-to-strong correlations with testosterone in young Tsimané males. Positive intercorrelations between measures of body masculinity and body size have also been reported in other studies (Acar & Eler, 2018; Apicella, 2014; Gallup et al., 2007; Shoup & Gallup, 2008; Sneade & Furnham, 2016; van Dongen & Sprengers, 2012; but see also Weeden & Sabini, 2007). While extant evidence therefore does appear to support the notion that masculine trait expression in men's faces and bodies may be related, intercorrelations with other masculine traits are more tentative. Voice pitch shows mixed relationships with other masculine traits (Aung & Puts, 2020; Cartei et al., 2014; Pisanski et al., 2014), and 2D:4D generally does not seem to correlate with other testosterone-dependent traits (Evans et al., 2008; Gallup et al., 2007; Lu et al., 2017; Neave et al., 2003; Rahman et al., 2005; Sim & Chun, 2016; but see also Acar & Eler, 2018; Fink et al., 2006).

1.4. Sexual selection models of how masculine traits evolved in human males

To reiterate, hormone-induced accelerated trait development in puberty is characteristic of a sexually selected trait. With the exception of 2D:4D, the masculine traits described above all appear to show the pubertal growth spurts one would expect of traits shaped by sexual selection. This suggests that sexual selection pressures might underlie masculine trait evolution in human males. However, in the current literature it is variously hypothesised that such pressures reflect female choice for 'good genes', signalled by increased masculinity, versus that masculine traits increase men's same-sex competitiveness for mates, resources, or social capital. The two most prevalent hypotheses proposed to explain masculine trait evolution in humans, the *immunocompetence handicap hypothesis* (Folstad & Karter, 1992) and the *male-male competition hypothesis* (Puts, 2016), are reviewed in detail in Chapter 2 (p.29-32); for brevity, they will not be discussed further here.

1.5. Predictions based on sexual selection models

Crucially, sexual selection models proposed to explain masculine trait evolution suggest that more masculine men have historically achieved greater biological fitness. While this may appear like a straightforward prediction, there is a number of pathways through which this might be achieved, but some of these have rarely been tested empirically. For example, masculine men might outcompete other men for mates; women might preferentially select masculine men; masculine men might expend more energy on mating effort; or masculine men might sire better-quality offspring who are more likely to survive.

This thesis will test three key predictions, derived from sexual selection models, across four empirical chapters: *i.* that sexual dimorphism positively predicts biological fitness in men and women; *ii.* that more dimorphic parents have better quality offspring; and *iii.* that more dimorphic (masculine) men have better quality partners. The first two chapters consider whether dimorphism directly influences reproductive success in men and women. In Chapter 2, we use a meta-analysis of existing published and unpublished data to test whether men's masculine traits increase their biological fitness (reproductive outcomes and mating proxies of reproduction). *If* more masculine men achieve greater fitness, one pathway through which that could be achieved is by mating with more fertile, better-quality partners. Similarly, in women, exaggerated feminine/attractive traits are commonly argued to index fertility. In Chapter 3, we therefore use both a systematic review and meta-analysis of the literature, and novel analyses of archival datasets to examine whether femininity/attractiveness indexes biological fitness in women; a commonly assumed but rarely tested prediction.

Chapter 4 focusses on offspring quality by investigating this as a function of parental dimorphism in two samples, to test a key prediction of the immunocompetence handicap hypothesis: that dimorphic traits in men indicate underlying heritable health. We also consider whether general parental attractiveness can predict heritable offspring health outcomes as well as earlier sexual activity (a proxy for mating success) in offspring.

Lastly, in Chapter 5, we return to the question of whether more masculine men have better quality partners, as well as whether they are of better quality themselves. If masculinity indexes men's quality (be it 'good genes' or increased competitiveness), it can be expected that women should find more masculine men attractive. The extent to which masculine traits in men are considered attractive by heterosexual women has garnered enormous attention in the literature, with particular focus on masculinity in men's faces (e.g., Alharbi et al., 2020; DeBruine et al., 2010; Docherty et al., 2020; Marcinkowska et a., 2018; 2019; 2021; Swaddle & Reierson, 2002). The extant literature on this question is beyond the scope of this thesis and will not be reviewed here. In summary, the evidence for masculine traits increasing men's attractiveness is tentative at best (this is particularly the case for facial masculinity), but this may depend on the trait being studied (e.g., Jones et al., 2010; Sell et al., 2017; Suire et al., 2019). If, however, masculinised men are more desirable mating or reproductive partners, we may find that they are able to attract higher quality partners of their own.

Chapter 2.

A meta-analysis of the association between male dimorphism and fitness outcomes in humans

---

2.1. Introduction

2.1.1. Sexual dimorphism and masculinity in humans

Sexual dimorphism refers to sex differences in morphological and behavioural traits, excluding reproductive organs (Plavcan, 2001), with particular emphasis on traits thought to have evolved through sexual selection (Crook, 2017). Humans are a sexually dimorphic species (Plavcan, 2001). Sexual selection in mammalian species, including human and non-human primates, is commonly argued to have acted more strongly on male traits, as a consequence of greater variance in males' reproductive output (Hammer et al., 2008) and a male-biased operational sex ratio, i.e., a surplus of reproductively available males relative to fertile females (e.g., Mitani et al., 1996).

Dimorphic traits that are exaggerated in males are typically referred to as masculine. In humans, masculine faces are characterised by features such as a pronounced brow ridge, a longer lower face, and wider mandibles, cheekbones, and chins (Swaddle & Reierson, 2002). Men are, on average, 7-8% taller than women (Gray & Wolfe, 1980) and weigh approximately 15% more (Smith & Jungers, 1997). Relative to this fairly modest body size dimorphism, upper body musculature and strength are highly dimorphic in humans: compared to women, men have 61% more overall muscle mass, and 90% greater upper body strength (Lassek & Gaulin, 2009). Men's bodies also tend to have a V- or wedge-shape, showing a greater shoulder-to-hip ratio (Hughes & Gallup, 2003; Singh, 1993) and waist-to-chest ratio (Tovée et al., 1999; Weeden & Sabini, 2007) than women's. Second-to-fourth finger (digit) length ratios are often claimed to be sexually dimorphic, with men's 2D:4D typically being lower than women's (Manning, 2002; though this may not be universal: Apicella et al., 2016). In addition, fundamental frequency, commonly referred to as voice pitch, is nearly six standard deviations lower in men than in women (Puts et al., 2012).

The development of these masculine traits in men is influenced by exposure to androgens, particularly testosterone. With the exception of 2D:4D, which is commonly claimed to be influenced primarily by prenatal testosterone levels and is present at birth (Galis et al., 2010; but see Richards et al., 2019), masculine traits generally develop or become exaggerated following a surge in testosterone production at sexual maturity (Butterfield et al., 2009; Fechner, 2003; Weston et al., 2003) – although it is not necessarily clear whether the size of that surge corresponds directly to the extent of trait expression.

2.1.2. Proposed mechanisms underlying the evolution of masculine traits

Key to the assumption that men's masculine traits are sexually selected is that masculine traits should be reliably associated with greater biological fitness. Men may increase fitness by producing a greater quantity of offspring overall (i.e., greater *fertility*), by acquiring a greater number of partners which may in turn mediate offspring numbers (greater *mating success*), and/or by producing more *surviving* offspring (greater *reproductive success*).

Two key hypotheses and attendant mechanisms have been drawn on by evolutionary behavioural scientists, predicting positive associations between masculinity and fitness outcomes. Firstly, according to the *immunocompetence handicap hypothesis* (Folstad & Karter, 1992), masculine traits are a costly signal of heritable immunocompetence, i.e., good genetic quality, due to the putative immunosuppressive properties of testosterone (see

Muehlenbein & Bribiescas, 2005). Masculine men should therefore produce healthier and more viable offspring, who are more likely to survive. Thus, women should be able to increase their fitness (via offspring survival) by selecting masculine men as mates. Authors therefore suggested that masculinity in men is intersexually selected, evolved and/or maintained through female choice, and should be associated with greater mating success in contexts where women are able to exercise choice. This should thus result in greater reproductive success, and an advantage in offspring survival.

The immunocompetence handicap hypothesis has persisted in the literature, particularly with reference to facial masculinity (although there are no *a priori* reasons to expect this putative mechanism to act more strongly on men’s faces than on their bodies), despite concerns regarding its validity since at least 2005 (Boothroyd et al., 2005). While beyond the scope of this article, common criticisms include that the relationship between testosterone and health is complex (Nowak et al., 2018), and facial masculinity is inconsistently linked to health (e.g., Boothroyd et al., 2013; Foo et al., 2020; Marcinkowska et al., 2019; Scott et al., 2013; Zaidi et al., 2019). Evidence is similarly mixed regarding the key assumption that women are attracted to masculinity in men’s faces (Boothroyd et al., 2013; Little, 2015) and bodies (Frederick & Haselton, 2007; Gray & Frederick, 2012; Lukaszewsi et al., 2014).

Secondly, under the *male-male competition hypothesis,* authors have argued that formidable (i.e., physically strong and imposing) men are better equipped to compete with other men for resources, status, and partners (Hill et al., 2017; Puts, 2016), through e.g., direct physical contests or by deterring rivals indirectly (Hill et al., 2017; Sell et al., 2012). For instance, increased musculature may intimidate competitors by signalling fighting prowess (Sell et al., 2009) and strength (Durkee et al., 2018), while facial masculinity and voice pitch may also have an indirect relationship with perceived formidability (Butovskaya et al., 2018;

Haselhuhn et al., 2015; Raine et al., 2018; Little et al., 2015; Puts & Aung, 2019; Scott et al., 2014).

Importantly, while male-male competition is often framed as an alternative to female choice, women may preferentially mate with both well-resourced men, and with competitive men, facilitating intersexual selection for masculinity (i.e., a ‘sexy sons’ effect, see Weatherhead & Robertson, 1979) where male status is due to, or competitiveness is cued by, formidability (Scott et al., 2013). Some authors have suggested that formidability increases men’s mating success through dominance over other men (which may create the circumstances that women select them as mates) rather than women’s direct preferences for formidable traits per se (Hill et al., 2013; Kordsmeyer et al., 2018; Slatcher et al., 2011). However, regardless of whether the driving mechanism is intra- or intersexual selection (or a combination thereof), the male-male competition hypothesis predicts that formidable men will acquire more partners over their lifetime, which will in turn result in more offspring. This approach, however, does not make any particular predictions regarding offspring health or survival.

It can be noted that proponents of both the immunocompetence and male-male competition hypotheses have also suggested that more masculine men may show reduced investment in romantic relationships and in offspring (Booth & Dabbs Jr, 1993; Boothroyd et al., 2007; Muller et al., 2009; Schild et al., 2020), potentially suppressing offspring health/survival. This could arise from an association between circulating testosterone (which masculine traits are commonly argued to index) and motivation for sexual behaviour (Grebe et al., 2019; Halpern et al., 1993) shifting effort away from parental investment towards pursuit of mating opportunities. Two important caveats here, however, are that the relationship between men’s testosterone levels in adolescence (when most masculine traits become exaggerated) and in adulthood is exceedingly weak (van Bokhoven et al., 2006), and masculine trait expression in adulthood is not consistently correlated with adult testosterone levels (e.g., Lefevre et al., 2013; Peters et al., 2008). Simply being more attractive to potential new partners, however, might shift behaviour away from relationship investment (for discussion see e.g., Gangestad & Simpson, 2000). Because of this, many authors have previously suggested that women face a trade-off between the (health or competitive) benefits of masculinity, and paternal investment.

2.1.3. The association between masculine traits and biological fitness

We therefore have at least two theoretical positions which assert that masculine men should have greater numbers of sexual partners, greater offspring numbers, and perhaps a greater proportion of surviving offspring, in at least some circumstances. Studies addressing these predictions in societies without effective contraception have done so directly via offspring numbers and/or offspring survival. In most industrialised populations, where access to contraceptives attenuates the relationship between sexual behaviour and reproductive success, mating success measures are often used instead. These include preferences for casual sex, number of sexual partners, and age at first sexual intercourse (earlier sexual activity allows for a greater lifetime number of sexual partners), as these are assumed to have correlated with reproductive success in men under ancestral conditions (Pérusse, 1993).

A key problem, however, is that the predictions outlined above do not always capture the diversity of human reproductive ecologies even where diverse data exists. We have already noted the fact that female choice may be important to outcomes above. Furthermore, even amongst non-contracepting populations, differences in rates of polygyny, pair-bond breakdown, and attitudes to fertility may moderate reproductive success and its variance. For instance, monogamous cultures do not typically show greater variance in men’s versus women’s reproductive success (Brown et al., 2009) and while increasing numbers of sexual partners (e.g., in serially monogamous or polygynous cultures) may often be important for increasing male reproductive success, the inverse is true amongst the Pimbwe where women are more advantaged by increased numbers of partners (Borgerhoff Mulder, 2009). Similarly, although the strongly monogamous Agta show high rates of fertility (Boothroyd et al., 2017), data from ostensibly non-contracepting rural Catholics in C20th Poland (Pawlowski et al., 2008) shows much lower rates of fertility. These issues highlight the fact that humans have likely had diverse reproductive and pair-bonding norms for a long time. As such we can make two observations. Firstly, availability of contraception in low-fertility samples might ‘free’ sexual behaviour from the constraints of pregnancy avoidance, and we might find *stronger* relationships between any evolved motivation for sex, and actual sexual behaviour, in these samples than would have necessarily been found ancestrally. Secondly, however, any adaptation which has been maintained across recent hominid lineages must have been adaptive *on average* across diverse reproductive ecologies. As such, if the proposed adaptation (masculinity leading to enhanced reproductive success via mating, and possibly increased offspring survival) exists, we should expect to see both: *i.* masculinity being associated with increased mating success in both high and (perhaps especially) low fertility populations, and *ii.* masculinity being on average positively associated with fertility, and potentially offspring survival, in non-contraception/high fertility populations.

2.1.4. Meta-analysis in sexual selection

Meta-analysis can be a valuable tool in understanding overall patterns in evolutionarily relevant traits, both across and within species. Jennions and colleagues (2012) noted that many traits hypothesised to predict male mating success had not been subject to meta-analysis, and further argued that while such meta-analyses can be valuable in clarifying the nature and extent of selection for some traits, at other times they act to refute prior assumptions. They say: “A general insight from sexual selection meta-analyses … is that it is easy to be misled by a few high-profile studies into believing that a prediction is well supported. Support is often weaker than assumed." (p.1139). This point does not just apply to comparative research, but is relevant to human sexual selection work specifically. For instance, van Dongen and Gangestad (2011) found that evidence for health benefits of symmetry were weaker and harder to demonstrate meta-analytically than they would have supposed, given the size of the extant literature. Similarly, when two meta-analyses into the effects of menstrual cycle on women's behaviour, mate preferences, and attractiveness reached opposing conclusions (Gildersleeve et al., 2014a; Wood et al., 2014), the exercise suggested that some cycle effects were unlikely to be robust. Indeed, the more cautious analytical methods (e.g., treating unknown null results as zero rather than excluding them from analysis) resulted in a null overall effect – a finding that was later borne out by multiple large, pre-registered, studies (Jones et al., 2018; Jünger et al., 2018; Marcinkowska et al., 2018). The authors of the meta-analysis that found a null effect suggested that publication and inclusion bias was a particular problem in the field (Harris et al., 2014), although others argued against this (Gildersleeve et al., 2014b).

In terms of the current topic, previous studies explicitly testing the relationships between masculine traits and fitness outcomes have been overwhelmingly conducted in low fertility samples and have produced a mixture of positive, negative, and null results (e.g., Boothroyd et al., 2017; Arnocky et al., 2018; Rhodes et al., 2005). This creates a clear need for meta-analytic comparison of evidence from as wide a population sample as possible. To date, however, meta-analytic analyses are rare, typically exclude many aspects of masculinity, and focus on *either* mating *or* reproductive outcomes, despite both being relevant to testing the theories above. Van Dongen and Sprengers (2012) meta-analysed the relationships between men's handgrip strength (HGS) and sexual behaviour in only three industrialised populations (showing a weak, positive association [$r = .24$]). Across 33 non-industrialised societies, von Rueden and Jaeggi (2016) found that male status (which included, but was not limited to, measures of height and strength) weakly predicted reproductive success (overall $r = .19$). In contrast, Xu and colleagues (2018) reported no significant association between men's height and offspring numbers across 16 studies when analysing both industrialised and non-industrialised populations. Lastly, Grebe and colleagues' (2019) meta-analysis of 16 effects – the majority of which came from Western samples - showed that men with high levels of circulating testosterone, assayed by blood or saliva, invested more in mating effort, indexed by mating with more partners and showing greater interest in casual sex ($r = .22$). Across all of their analyses (which also included pair-bond status, fatherhood status, and fathering behaviours), Grebe and colleagues found no significant differences between 'Western' and 'non-Western' samples, but their 'non-Western' grouping for the relevant analysis only included a low fertility population in 21st Century China. To our knowledge, facial masculinity, voice pitch, and 2D:4D have never been meta-analyzed in relation to mating and/or reproduction.

2.1.5. The present study

The present article (Study 1) therefore searched widely for published and unpublished data to meta-analyse the relationships between six main masculine traits in men (facial masculinity, body masculinity, 2D:4D, voice pitch, height, and testosterone levels) and *both* mating and reproductive outcomes, in *both* high and low fertility samples. By including multiple traits, a broad search strategy, and considering high and low fertility samples both separately and together, we can ascertain whether the current scientific evidence base provides plausible support for the sexual selection of masculine traits in humans. By further testing the publication status of each effect (whether the specific effect size/analysis was reported in a published article or not), we can also evaluate the evidence for publication bias, since this is known to artificially inflate effects in diverse literatures.

Mating measures included behavioural measures such as number of sexual partners, number of marital spouses, and age at first sexual intercourse. Since increased mating effort is an additional possible route to increased reproductive output, we also included mating attitudes, such as preferences for casual sex. Reproductive measures included: fertility measures, such as number of children/grandchildren born and age at the birth of the first child; and reproductive success measures, i.e., number of offspring surviving childhood. Since offspring mortality is a measure specifically of offspring viability, we included this as a separate measure (i.e., mortality rate and/or number of deceased offspring).

2.2. Methods

2.2.1. Literature search and study selection

A systematic search was initially carried out between November 2017 and February 2018 using the databases PsycINFO, PubMed, and Web of Science; the searches were saved and search alerts ensured inclusion of subsequently published studies. Search terms are given in Box 1.

Studies were also retrieved through cross-referencing, citation searches/alerts, and by asking for data on social media. The systematic search generated 2,221 results, including duplicates, and a further approximately 300 articles were found by other means. After scanning titles and abstracts, 280 articles/dissertations were reviewed in full. Studies submitted up to 1 May 2020 were accepted. Eligible studies included at least one of the following predictors: facial masculinity, body masculinity (strength, body shape, or muscle mass/non-fat body mass), 2D:4D, voice pitch, height, or testosterone levels.

Box 1.

Search terms for meta-analysis study discovery

(masculin OR "sexual dimorphism" OR "sexually dimorphic" OR width-to-height OR muscularity OR shoulder-to-hip OR chest-to-waist OR "digit ratio" OR 2d:4d OR "hand grip strength" OR "handgrip strength" OR "grip strength" OR testosterone OR "voice pitch" OR "vocal pitch" OR voice OR "non-fat body mass" OR "lean body mass" OR "fundamental frequency" OR "facial* dominan*" OR height OR "sexual dimorphism in stature" OR "CAG repeat*")*

AND

("sex partner*" OR "short-term relationship*" OR "short term mating" OR "extra pair" OR sociosexual* OR "age of first intercourse" OR "age of first sexual intercourse" OR "age at first intercourse" OR "age at first sexual intercourse" OR "age of sexual debut" OR "age at first sex" OR "mating success" OR "number of offspring" OR "offspring number" OR "number of children" OR "number of grandoffspring" OR "number of grand offspring" OR "offspring health" OR "offspring mortality" OR "mortality of offspring" OR "surviving offspring" OR "offspring survival" OR "reproductive onset" OR "reproductive success" OR "long-term relationship*" OR "age of first birth")*

AND (human OR man OR men OR participant).*

The following outcome measures were included:

- Mating domain: global sociosexuality (i.e., preferences for casual sex: Penke & Asendorpf, 2008; Simpson & Gangestad, 1991) and specific measures of mating attitudes and mating behaviours where:

i. Mating attitudes included: preferences for short-term relationships, and sociosexual attitudes and desires.

ii. Mating behaviours included: number of sexual partners, one-night-stands/short-term relationships, potential conceptions, sociosexual behaviours, extra-pair sex, age at first sexual intercourse, and number of marital spouses.

- Reproductive domain: including both fertility and reproductive success, described below.

i. Fertility: number of children and grandchildren born, and age at the birth of the first child.

ii. Reproductive success: number of surviving children/grandchildren.

- Offspring mortality domain: mortality rate and number of deceased offspring.

Both published and unpublished studies were eligible. We restricted our sample to studies with adult participants (≥17 years old). If key variables were collected but the relevant analyses were not reported, we contacted authors to request effect sizes or raw data. If data were reported in more than one study, we selected the analysis with the larger sample size or which included appropriate control variables, such as age. Studies using measures that were ambiguous and/or not comparable to measures used in other studies were excluded (e.g., measures of body size without information about the proportion of fat/muscle mass, or reproductive data during a very restricted time period). Twin studies where participants were sampled as pairs, population level studies, and studies analysing both sexes together were also excluded, as well as articles that were not written in English or Swedish as we were not sufficiently fluent in other languages to conduct unbiased searching and extraction. Multiple measures from the same study were retained if they met the other criteria.

We chose Pearson's *r* as our effect size measure and effect sizes not given as *r* were converted (see Appendix A for conversion formulas); if effect sizes were not convertible, the study was excluded. Where effect sizes for non-significant results were not stated in the article and could not be obtained, an effect size of 0 was assigned (*k* = 28). Excluding those effects from the analyses had no effect on any of the results. Twenty-nine percent of all observations (133 of 452, selected randomly) were double coded by the first author > 2 months apart. Intracoder agreement was 97%. For coding decisions, see Appendices B.

In total, 96 studies were selected (Alvergne et al., 2009; Apicella, 2014; Apicella et al., 2007; Arnocky et al., 2018; Aronoff, 2017; Atkinson, 2012; Atkinson et al., 2012; Bogaert & Fisher, 1995; Booth et al., 1999; Boothroyd et al., 2011; Boothroyd et al., 2008; Boothroyd et al., 2017; Charles & Alexander, 2011; Chaudhary et al., 2015; Edelstein et al., 2011; Falcon, 2016; Farrelly et al., 2015; Frederick, 2010; Frederick & Haselton, 2007; Frederick & Jenkins, 2015; Gallup et al., 2007; Genovese, 2008; Gettler et al., 2019; Gildner, 2018; Gómez-Valdés et al., 2013; Hartl et al., 2013; Hill et al., 2013; Hoppler et al., 2018; Hughes & Gallup, 2003; Hönekopp et al., 2006; Hönekopp et al., 2007; Kirchengast, 2000; Kirchengast & Winkler, 1995; Klimas et al., 2019; Klimek et al., 2014; Kordsmeyer & Penke, 2017; Kordsmeyer et al., 2018; Krzyżanowska et al., 2015; Kurzban & Weeden, 2005; Lassek & Gaulin, 2009; Little et al., 1989; Loehr & O'Hara, 2013; Longman et al., 2018; Luevano et al., 2018; Lukaszewski et al., 2014; Maestripieri et al., 2014; Manning & Fink, 2008; Manning et al., 2003; Marczak et al., 2018; McIntyre et al., 2006; Međedović & Bulut, 2019; Mosing et al., 2015; Muller & Mazur, 1997; Nagelkerke et al., 2006; Nettle, 2002; Pawlowski et al., 2000; Pawlowski et al., 2008; Peters et al., 2008; Pollet et al., 2011; Polo et al., 2019; Price et al., 2013; Prokop & Fedor, 2011; Prokop & Fedor, 2013; Puts et al., 2006; Puts et al., 2015; Putz et al., 2004; Rahman et al., 2005; Rhodes et al., 2005; Rosenfield et al., 2020; Schwarz et al., 2011; Scott & Bajema, 1982; Shoup & Gallup, 2008; Sim & Chun, 2016; Simmons & Roney, 2011; Smith et al., 2017; Sneade & Furnham, 2016; Sorokowski et al., 2013; Steiner, 2011; Stern et al., 2020; Strong, 2014; Strong & Luevano, 2014; Subramanian et al., 2009; Suire et al., 2018; Tao & Yin, 2016; van Anders et al., 2007; Van Dongen &

Sprengers, 2012; Varella et al., 2014; von Rueden et al., 2011; Voracek et al., 2010; Walther et al., 2016; Walther et al., 2017a; Walther et al., 2017b; Walther et al, 2017c; Waynforth, 1998; Winkler & Kirchengast, 1994; Weeden & Sabini, 2007), comprising 474 effect sizes from 99 samples and 177,044 unique participants (Appendix C). This exceeds the number of studies for each of the meta-analyses published previously (Grebe et al., 2019; Van Dongen & Sprengers, 2012; von Rueden & Jaeggi, 2016; Xu et al., 2018).

2.2.2. Statistical analyses

We used the *metafor* package (Viechtbauer, 2010) in R 3.6.2 (R Development Core Team, 2019). *metafor* transforms Pearson's *r* to Fisher's *Z* for analysis; for ease of interpretation, effect sizes were converted back to *r* for presentation of results. For 2D:4D and voice pitch, effects were reverse coded prior to analysis because low values denote greater masculinity. Similarly, effects were reverse coded for all offspring mortality outcomes as well as the outcomes age at first birth and age at first sexual intercourse/contact, as low values denote increased fitness. In all analyses reported here, therefore, a positive value of *r* denotes a positive relationship between masculinity and fitness outcomes. All predicted relationships were positive.

Analyses were conducted using random-effects models, as we expected the true effect to vary across samples. We controlled for multiple comparisons by computing *q*-values (Storey, 2002). Note that *q*-values estimate the probability that a significant effect is truly significant or not; they are not adjusted *p* values. Thus, in all analyses presented below, only effects that remained significant after *q*-value computation (indicated by *q*-values < .05) are presented as significant. We computed *q*-values using all *p* values across all tests conducted in the whole analysis (266 in total).

The analyses were conducted on three levels for both predictor traits and outcomes (Figure 2.1). For predictor traits, all six masculine traits were first combined and analysed together at the *global masculinity level*. At the *trait level*, each masculine trait was then analysed separately. Lastly, each masculine trait was further divided into separate *trait indices*, which were analysed as potential moderators (see below).

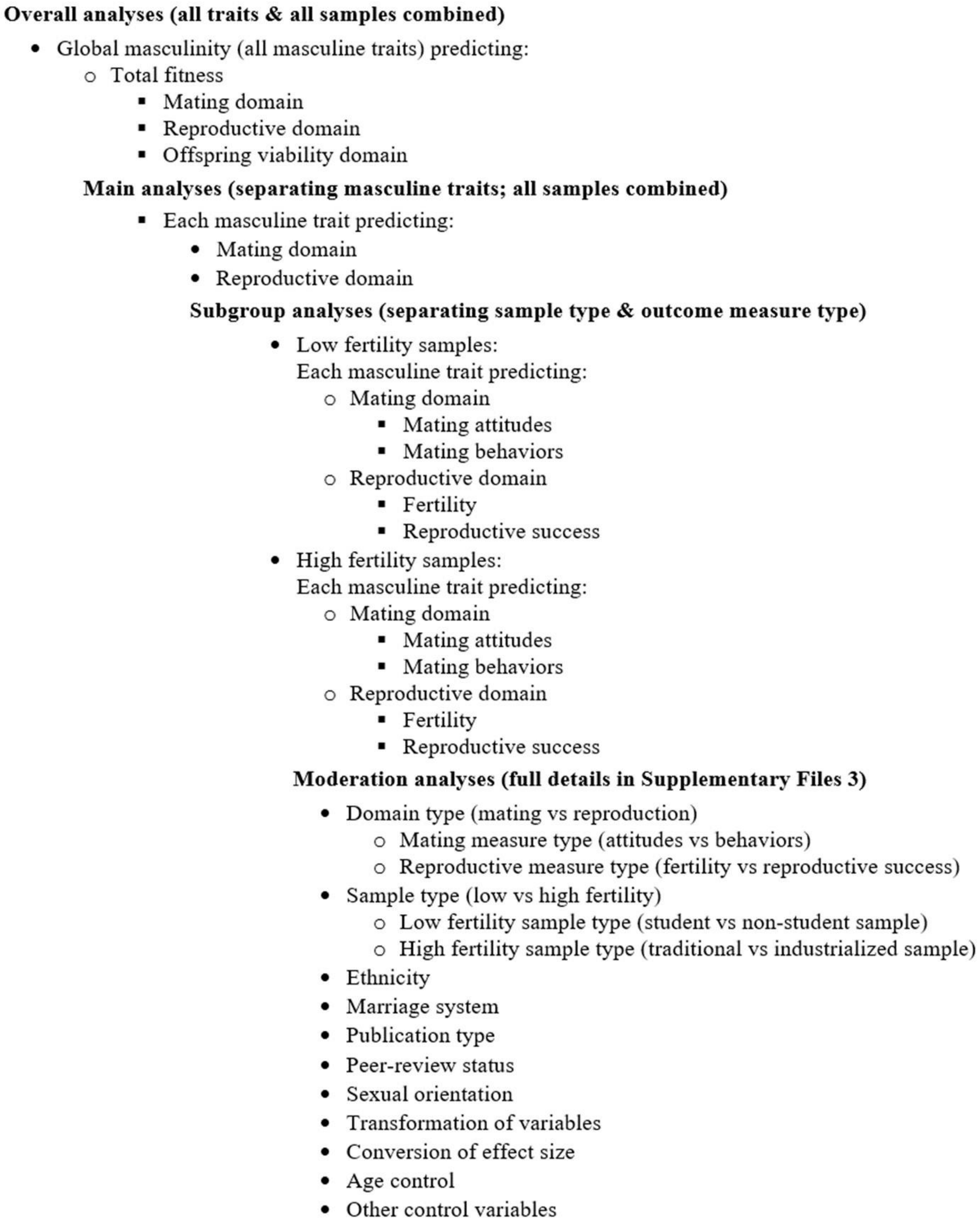
For the outcomes, mating, reproduction, and offspring mortality were first analysed together at the *total fitness level*. Given the widespread use of mating measures as proxies of reproductive outcomes, it is imperative where possible to test (and ideally compare) both mating and reproduction, to ensure that we are not relying on proxies that do not measure what they are assumed to measure. The *domain level* therefore divided outcomes into the *mating domain*, the *reproductive domain*, and the *offspring mortality domain* and analysed them separately. The last level, the *measures level*, further divided mating and reproduction into their separate measures (mating attitudes and behaviours, and fertility and reproductive success, respectively), which were analysed as subgroups.

The mating domain comprised mating *attitudes* and mating *behaviours,* as high mating success may result from increased mating efforts (reflected in favourable attitudes towards short-term mating) and/or encountering more mating opportunities (reflected in mating behaviours) without actively seeking them (because of female choice, for example). It is therefore necessary to divide these two measures.

The reproductive measures, fertility (number of offspring) and reproductive success (number of surviving offspring), are closely related but were also analysed separately in subgroup analyses. Offspring mortality, on the other hand, was usually indexed by mortality *rate* (only two studies used absolute number of dead offspring, and it made no difference to the results whether those studies were included or not) and is not directly related to offspring numbers. Offspring mortality was therefore analysed as a separate domain. As there were too

Figure 2.1

Overall analysis structure



Overall analyses (all traits & all samples combined)

- Global masculinity (all masculine traits) predicting:
    - Total fitness
        - Mating domain
        - Reproductive domain
        - Offspring viability domain

Main analyses (separating masculine traits; all samples combined)

- Each masculine trait predicting:
    - Mating domain
    - Reproductive domain

Subgroup analyses (separating sample type & outcome measure type)

- Low fertility samples:
  Each masculine trait predicting:
    - Mating domain
        - Mating attitudes
        - Mating behaviors
    - Reproductive domain
        - Fertility
        - Reproductive success
- High fertility samples:
  Each masculine trait predicting:
    - Mating domain
        - Mating attitudes
        - Mating behaviors
    - Reproductive domain
        - Fertility
        - Reproductive success

Moderation analyses (full details in Supplementary Files 3)

- Domain type (mating vs reproduction)
    - Mating measure type (attitudes vs behaviors)
    - Reproductive measure type (fertility vs reproductive success)
- Sample type (low vs high fertility)
    - Low fertility sample type (student vs non-student sample)
    - High fertility sample type (traditional vs industrialized sample)
- Ethnicity
- Marriage system
- Publication type
- Peer-review status
- Sexual orientation
- Transformation of variables
- Conversion of effect size
- Age control
- Other control variables

few observations of offspring mortality to test predictor traits separately, this outcome was only analysed at the global masculinity level.

In addition to analysing all samples together, we also analysed low and high fertility samples separately to assess whether results were robust in both types of populations. We used a cut-off of three or more children per woman on average within that sample, which roughly corresponds to samples with vs without widespread access to contraception (The World Bank, 2018). Samples therefore had two levels: *all samples*, and the two sample types *low fertility* and *high fertility*.

The analysis structure was therefore as summarised in Figure 2.1: *overall analyses* tested global masculinity as a predictor of total fitness, as well as the three domains of mating, reproduction, and offspring mortality, separately, across all samples. In our *main analyses*, we analysed masculinity at the trait level, in relation to the two outcome domains mating and reproduction. The following *subgroup analyses* considered low and high fertility samples separately, in addition to also dividing outcomes into their respective measures (mating attitudes vs mating behaviours, and fertility vs reproductive success).

Lastly, we performed a series of exploratory meta-regressions on potential moderator variables. Such moderation analyses compare effect sizes across categories of studies as determined by a particular study characteristic, e.g., monogamous vs polygynous marriage systems, to determine if effect sizes were robust and/or equivalent across these categories. Since power was often low, we ran moderation analyses separately for each study characteristic rather than trying to test for interactions. For all masculine traits where we had sufficient power, trait-general moderation analyses included: domain type (mating vs reproduction), mating measure type (attitudes vs behaviours), reproductive measure type (fertility vs reproductive success), sample type (low vs high fertility), low fertility sample type (student vs non-student), high fertility sample type (traditional vs industrialised), ethnicity, marriage system, publication status (published vs not published effect), peer review status (peer reviewed vs not peer reviewed), sexual orientation, transformation of variables, conversion of effect sizes, age control, and inclusion of other control variables. Note that since we included many non-published effects from studies that were published, ‘publication status' referred to whether particular the particular *effects* were published, not the study as whole. The analysis can therefore detect evidence of any tendency for significant results to be 'written up' while nonsignificant ones are not, whether this bias occurs between or within manuscripts. We ran moderation analyses both for outcome domains and outcome measures (i.e., mating attitudes and mating behaviours, and fertility and reproductive success, respectively). For each masculine trait, we also conducted trait-specific moderation analyses (e.g., subjectively rated vs morphometric facial masculinity (for full details on trait-specific moderators, see Appendices D).

Analyses sometimes included more than one observation from the same study/sample. In all analyses, therefore, effect sizes were clustered both by sample and by study. For all analyses, only relationships with a minimum of three independent samples from a minimum of two separate studies were analysed. For moderation analyses, this meant that *each category* of the moderator needed observations from at least three samples from at least two studies; in many cases, there were not enough observations to test for moderators.

In the Results section, unless otherwise specified, we summarise results from *trait-general* moderation analyses of outcome *domains* only (where results for outcome *measures* and *trait-specific* moderators are reported in Appendices E). Additional details and full results of all analyses can be found in Appendices D-F.

2.3. Results

2.3.1. Summary of samples

All 96 studies included in the meta-analysis are shown in Appendix C. In total, 29 articles reported effect sizes from high fertility samples, which included 17 articles drawing on 13 different extant forager or subsistence populations (of the type sometimes referred to as 'small scale societies', coded here as non-industrialised) predominantly in Africa or Latin America.

The remaining high fertility data came from historical samples or low socioeconomic status sub-populations within low-fertility countries (e.g., agricultural Polish communities, former 'delinquents' in the U.S., and Zulus living in South African townships). Sixty-nine articles reported data from low fertility populations, which came from 54 primarily student or partially-student samples (43 of which were from English-speaking countries), and only 12 samples which could be considered representative community or cohort/panel samples. Two articles reported data drawn from 'global' online samples (classified as low fertility). The remaining low fertility samples were either unspecified or sampled particular sub-populations (e.g., specific professions).

2.3.2. Overall analyses of global masculinity

In the initial overall analyses, global masculinity was weakly but significantly associated with greater total fitness (i.e., mating, reproduction, and offspring mortality combined) (r = .080, 95% CI: [0.061, 0.101], q = .001; we reiterate here that for all analyses, q-values < .05 denote significance after correcting for multiple comparisons). When we divided the outcome measures into their three domains, the positive (albeit weak) associations with global masculinity remained significant for mating, but not for reproduction or offspring mortality (mating: r = .090, 95% CI: [0.071, 0.110], q = .001; reproduction: r = .047, 95% CI: [0.004, 0.090], q = .080; offspring mortality: r = .002, 95% CI: [-0.011, 0.015], q = .475). While the effect was thus only significant for mating, the differences between effects were not significant, but we note that sample sizes differed considerably between domains.

Below, we present in further detail the results of the effect of global masculinity on each of the three outcome domains: mating, reproduction, and offspring mortality. We then present the associations between each masculine trait and mating and reproductive measures, separately. We also present results for subgroup and trait-general moderation analyses (for outcome domains only); for complete results, see Appendices E-F.

2.3.3. Mating

Main analyses of each masculine trait. This set of analyses tested the prediction that individual masculine traits are positively associated with mating. In terms of the overall mating domain (i.e., mating attitudes and behaviours combined), all masculine traits showed the predicted positive relationships with mating, and the effects were significant for all traits except for facial masculinity and 2D:4D (Table 2.1). Some of these effects were very weak, however. The strongest associations with the mating domain were seen in terms of body masculinity (r = .133, 95% CI: [0.091, 0.176], q = .001; Appendix G.1), voice pitch (r = .132, 95% CI: [0.061, 0.204], q = .002; Appendix G.2), and testosterone levels (r = .093, 95% CI: [0.066, 0.121], q = .001; Appendix G.3). Height showed a significant but smaller effect size (r = .057, 95% CI: [0.027, 0.087], q = .002; Appendix G.4). While not the weakest association, the relationship between facial masculinity and mating was nonsignificant (r = .080, 95% CI: [-0.003, 0.164], q = .117). The effect for 2D:4D was also nonsignificant (r = .034, 95% CI: [0.000, 0.069], q = .102), and moderation analyses showed that this was the only trait that showed a significantly smaller effect size than the strongest predictor, body masculinity (p < .001, q = .006).

Table 2.1

Masculine traits predicting mating: main analyses and subgroup analyses of mating attitudes vs mating behaviours and low vs high fertility samples. Pearson's r (95% CI); p value for meta-analytic effect, q-value (correcting for multiple comparisons); number of observations (k), samples (s), and unique participants (n); test for heterogeneity (Q), p value for heterogeneity. Statistically significant meta-analytic associations are bolded if still significant after controlling for multiple comparisons.

	Mating		
Outcome: Sample	Facial masculinity	Body masculinity	2D:4D
Mating domain: All samples	*r* = .080 (-0.003, 0.164), *p* = .060, *q* = .117 *k* = 30, *s* = 11, *n* = 948 Q(df = 29) = 54.834, *p* = .003	***r* = .133 (0.091, 0.176)**, ***p* < .001, *q* = .001** *k* = 121, *s* = 32, *n* = 7939 Q(df = 120) = 297.472, *p* < .001	*r* = .034 (0.000, 0.069), *p* = .049, *q* = .102 k = 84, s = 23, n = 66807 Q(df = 83) = 101.994, *p* = .077
Mating att: All samples	*r* = .095 (-0.072, 0.263), *p* = .263, *q* = .304 *k* = 5, *s* = 4, *n* = 407 Q(df = 4) = 8.684, *p* = .070	*r* = .078 (0.002, 0.155), *p* = .045, *q* = .098 *k* = 20, *s* = 9, *n* = 922 Q(df = 19) = 17.606, *p* = .549	*r* = .035 (-0.061, 0.132), *p* = .474, *q* = .385 *k* = 19, *s* = 7, *n* = 504 Q(df = 18) = 24.141, *p* = .151
Mating beh: All samples	*r* = .025 (-0.059, 0.109), *p* = .554, *q* = .424 *k* = 22, *s* = 8, *n* = 755 Q(df = 21) = 37.044, *p* = .017	***r* = .142 (0.099, 0.187)**, ***p* < .001, *q* = .001** *k* = 91, *s* = 31, *n* = 7738 Q(df = 90) = 267.876, *p* < .001	*r* = .038 (-0.002, 0.078), *p* = .061, *q* = .117 *k* = 51, *s* = 19, *n* = 1607 Q(df = 50) = 64.049, *p* = .087
Mating domain: Low fert. samples	*r* = .089 (-0.001, 0.179), *p* = .053, *q* = .109 *k* = 28, *s* = 10, *n* = 913 Q(df = 27) = 54.287, *p* = .001	***r* = .135 (0.091, 0.180)**, ***p* < .001 , *q* = .001** *k* = 117, *s* = 28, *n* = 7572 Q(df = 116) = 289.080, *p* < .001	*r* = .038 (0.002, 0.073), *p* = .037, *q* = .086 *k* = 82, *s* = 22, *n* = 66751 Q(df = 81) = 101.369, *p* = .063
Mating att: Low fert. samples	*r* = .095 (-0.072, 0.262), *p* = .263, *q* = .304 *k* = 5, *s* = 4, *n* = 407 Q(df = 4) = 8.684, *p* = .070	*r* = .078 (0.002, 0.155), *p* = .045, *q* = .098 *k* = 20, *s* = 9, *n* = 922 Q(df = 19) = 17.606, *p* = .549	*r* = .035 (-0.061, 0.132), *p* = .474, *q* = .385 *k* = 19, *s* = 7, *n* = 504 Q(df = 18) = 24.141, *p* = .151
Mating beh: Low fert. samples	*r* = .028 (-0.063, 0.119), *p* = .543, *q* = .420 *k* = 20, *s* = 7, *n* = 720 Q(df = 19) = 36.610, *p* = .009	***r* = .145 (0.100, 0.193)**, ***p* < .001, *q* = .001** *k* = 87, *s* = 27, *n* = 7371 Q(df = 86) = 259.448, *p* < .001	*r* = .042 (0.001, 0.083), *p* = .045, *q* = .098 *k* = 49, *s* = 19, *n* = 1551 Q(df = 48) = 62.941, *p* = .073
Mating domain: High fert. samples	*s* = 1	*r* = .105 (-0.069, 0.280), *p* = .235, *q* = .285 *k* = 4, *s* = 4, *n* = 367 Q(df = 3) = 7.282, *p* = .063	*s* = 1

Mating att: High fert. samples	s = 0	s = 0	s = 0
Mating beh: High fert. samples	s = 1	r = .105 (-0.069, 0.280), p = .235, q = .285 k = 4, s = 4, n = 367 Q(df = 3) = 7.282, p = .063	s = 1

	Voice pitch	Height	T levels
Mating domain: All samples	**r = .132 (0.061, 0.204)**, **p < .001, q = .002** k = 8, s = 5, n = 443 Q(df = 7) = 2.334, p = .939	**r = .057 (0.027, 0.087)**, **p < .001, q = .002** k = 62, s = 25, n = 43686 Q(df = 61) = 263.247, p < .001	**r = .093 (0.066, 0.121)**, **p < .001, q = .001** k = 66, s = 21, n = 7083 Q(df = 65) = 66.090, p = .439
Mating att: All samples	s = 0	r = .028 (-0.013, 0.068), p = .179, q = .253 k = 9, s = 6, n = 4232 Q(df = 8) = 5.137, p = .743	**r = .099 (0.026, 0.173)**, **p = .008, q = .032** k = 21, s = 11, n = 1039 Q(df = 20) = 25.379, p = .187
Mating beh: All samples	**r = .124 (0.043, 0.206)**, **p = .003, q = .016** k = 7, s = 5, n = 443 Q(df = 6) = 2.162, p = .904	**r = .054 (0.021, 0.087)**, **p = .001, q = .008** k = 48, s = 24, n = 42179 Q(df = 47) = 247.032, p < .001	**r = .084 (0.058, 0.110)**, **p < .001, q = .001** k = 32, s = 17, n = 6765 Q(df = 31) = 28.558, p = .592
Mating domain: Low fert. samples	**r = .129 (0.055, 0.204)**, **p < .001, q = .005** k = 7, s = 4, n = 388 Q(df = 6) = 2.234, p = .897	**r = .055 (0.024, 0.086)**, **p < .001, q = .004** k = 58, s = 21, n = 43310 Q(df = 57) = 259.576, p < .001	**r = .099 (0.069, 0.129)**, **p < .001, q = .001** k = 58, s = 20, n = 6795 Q(df = 57) = 61.443, p = .320
Mating att: Low fert. samples	s = 0	r = .028 (-0.013, 0.068), p = .179, q = .253 k = 9, s = 6, n = 4232 Q(df = 8) = 5.137, p = .743	**r = .108 (0.021, 0.195)**, **p = .015, q = .047** k = 17, s = 10, n = 751 Q(df = 16) = 20.017, p = .220
Mating beh: Low fert. samples	**r = .119 (0.034, 0.205)**, **p = .006, q = .025** k = 6, s = 4, n = 388 Q(df = 5) = 2.017, p = .847	**r = .051 (0.017, 0.086)**, **p = .004, q = .019** k = 44, s = 20, n = 41803 Q(df = 43) = 243.392, p < .001	**r = .088 (0.058, 0.119), p < .001, q = .001** k = 30, s = 16, n = 6477 Q(df = 29) = 27.793, p = .529
Mating domain: High fert. samples	s = 1	r = .089 (-0.016, 0.193), p = .096, q = .157 k = 4, s = 4, n = 376 Q(df = 3) = 3.388, p = .336	s = 1
Mating att:	s = 0	s = 0	s = 1

High fert. samples			
Mating beh: High fert. samples	*s* = 1	*r* = .089 (-0.016, 0.193), *p* = .096, *q* = .157 *k* = 4, *s* = 4, *n* = 376 Q(df = 3) = 3.388, *p* = .336	*s* = 1

Note. Att = attitudes; beh = behaviours; fert = fertility; *k* = number of observations; *n* = number of unique participants; Q = Cochran's Q test of heterogeneity; *q* = *q*-value; *s* = number of samples; T = testosterone.

Comparison of high and low fertility samples. Across all masculine traits, most effect sizes (94%) came from low fertility samples. Moderation analyses of sample type could only be run for body masculinity and height; neither was significant, although in both cases the effect sizes observed in the main analyses were significant only for low fertility, and not the less numerous high fertility samples (*k* = 4 for each trait). The other four traits had only been measured in one high fertility sample each, and the main analyses thus contained almost exclusively low fertility samples. We further compared low fertility samples which were predominantly students with other low fertility samples as part of our moderation analyses where possible, i.e., for body masculinity, voice pitch, height, and testosterone. For body masculinity, student samples showed a significantly stronger effect than non-student samples for mating *behaviours* only (*B* = -.128, *p* = .009, *q* = .032) but otherwise we found no differences (see Appendices E).

Inclusion bias/heterogeneity. Since the analysis included unpublished data, the distribution of effects in the funnel plots (see Appendix H.1) shows availability bias rather than publication bias. Apart from voice pitch, for which we did not have many effects, visual inspection of funnel plots indicated that they were generally symmetric, suggesting that the analysis did not systematically lack studies with unexpected small effects. There was significant heterogeneity of effect sizes for facial masculinity, body masculinity, and height; all of which are accounted for in a random-effects analysis.

Additional subgroup and moderation analyses for outcome domains. In this step of the analyses, we tested the hypothesis that each of the six masculine traits is positively associated with the two mating domain measures (mating attitudes and mating behaviours) and tested further potential control variables and trait-specific moderators. Results of subgroup analyses can be viewed in Table 2.1 and trait-general moderators in Table 2.3; full results of all moderation analyses are reported in Appendices E.

Table 2.3

Overview of moderation analyses for the mating vs reproductive domains. Significant associations are indicated by + and – signs, showing the direction of the moderator relative to the reference category (stated first in the moderator column); crosses indicate no significant moderation; and 'na' indicates that power was too low to run that specific analysis. Only associations that remained significant after controlling for multiple comparisons are indicated here.

Note that this table only shows general moderators shared by all masculine traits; for trait-specific moderation analyses, see Appendices E. Likewise, for moderation analyses of the two mating domain measures attitudes and behaviours, and the two reproductive domain measures fertility and reproductive success, we also refer to Appendices E.

	Facial masc.		Body masc.		2D:4D		Voice pitch		Height		T levels	
Moderator	MAT	REP	MAT	REP	MAT	REP	MAT	REP	MAT	REP	MAT	REP
Mating vs reproductive domain	✗		✗		✗		✗		✗		✗	
Mating attitudes vs behaviours	✗	na	✗	na	✗	na	na	na	✗	na	✗	na
Fertility vs reproductive success	na	na	na	✗	na	✗	na	na	na	✗	na	na
Low vs high fertility sample	na	na	✗	na	na	✗	na	na	✗	✗	na	na
Low fertility: student vs non-student sample	na	na	✗	na	✗	na	na	na	✗	na	✗	na
High fertility: traditional vs industrialised sample	na	na	na	na	na	✗	na	na	na	✗	na	na
Predominantly white vs mixed/other/unknown ethnicity sample	✗	na	✗	na	–	✗	na	na	✗	✗	✗	na
Monogamous vs non-monogamous marriage system	na	na	✗	na	na	✗	na	na	na	✗	na	na

Published vs non-published results	×	na	×	×	×	×	na	na	×	×	×	na
Peer reviewed vs not peer reviewed study	na	na	×	na	×	na	na	na	×	na	na	na
Heterosexual vs gay/mixed/ unknown sample	×	na	×	na	×	×	na	na	×	+	−	na
Non-normality-transformed vs transformed variables	na	na	×	×	+	×	na	na	×	×	+	na
Non-converted vs converted effect sizes	na	na	×	+	na	na	na	na	×	×	×	na
Age controlled for vs not controlled for	×	na	+	na	×	×	na	na	×	×	×	na
Inclusion of non-relevant control variables vs not	na	na	na	×	na	na	na	na	na	×	×	na

Note. Masc = masculinity; MAT = mating; REP = reproduction; T = testosterone.

Type of mating measure (attitudes vs behaviours) was never a significant moderator. However, for both body masculinity and height, there were significant effects for mating behaviours (body masculinity: r = .142, 95% CI: [0.099, 0.187], q = .001, height: r = .054, 95% CI: [0.021, 0.087], q = .008) but not attitudes. Voice pitch was significantly related to mating behaviours (r = .124, 95% CI: [0.043, 0.206], q = .016) but was not measured in combination with mating attitudes. Testosterone levels showed near identical effects for both mating attitudes and behaviours (r = .099, 95% CI: [0.026, 0.173], q = .032 and r = .084, 95% CI: [0.058, 0.110], q = .001, respectively).

No trait-general moderator consistently changed the pattern of the associations (Table 2.3). Body masculinity effects were stronger in studies where age had not been controlled for compared to where it had been controlled for (B = 0.103, p = .015, q = .047). Associations for 2D:4D were weaker in non-white/mixed ethnicity samples compared to white samples (B = -0.080, p = .014, q = .047), and stronger where variables had been transformed to approximate normality compared to when they had not been transformed (B = 0.103, p = .016, q = .047).

Similarly, associations for testosterone levels were also stronger for normality-transformed variables (*B* = 0.057, *p* = .015, *q* = .047), and weaker in gay/mixed sexuality samples compared to in heterosexual samples (*B* = -0.059, *p* = .003, *q* = .016).

For trait-specific moderators, significant moderation was seen for *type* of body masculinity where body shape was a significantly weaker predictor than strength (*B* = -0.099, *p* = .003, *q* = .017). Effects for *rated* body masculinity were significantly stronger than for indices taken from body measurements (*B* = 0.177, *p* = .007, *q* = .029). For 2D:4D, studies that had measured digit ratios three times – rather than twice or an unknown number of times – showed significantly stronger effects (*B* = 0.102, *p* = .006, *q* = .025).

2.3.4. Reproduction

Main analyses of each masculine trait. In this set of analyses, we tested the hypothesis that individual masculine traits positively predict reproduction. As Table 2.2 shows, relationships were generally in the predicted direction, but body masculinity was the strongest and only significant predictor (*r* = .143, 95% CI: [0.076, 0.212], *q* = .001; Figure 2.2). The only trait with an effect size significantly smaller than body masculinity was height (*B* = -0.107, *p* = .005, *q* = .023).

Table 2.2

Masculine traits predicting reproduction: main analyses and subgroup analyses of fertility vs reproductive success and low vs high fertility samples. Pearson's r (95% CI); p value for meta-analytic effect, q-value (correcting for multiple comparisons); number of observations (k), samples (s), and unique participants (n); test for heterogeneity (Q), p value for heterogeneity. Statistically significant meta-analytic associations are bolded if still significant after controlling for multiple comparisons.

Outcome: Sample	Reproduction		
	Facial masculinity	Body masculinity	2D:4D
Repr. domain: All samples	*r* = .099 (-0.012, 0.211), *p* = .081, *q* = .140 *k* = 5, *s* = 5, *n* = 1411 Q(df = 4) = 8.799,	***r* = .143 (0.076, 0.212)**, ***p* < .001, *q* = .001** *k* = 14, *s* = 8, *n* = 897 Q(df = 13) = 16.356,	*r* = .074 (-0.006, 0.154), *p* = .070, *q* = .131 *k* = 19, *s* = 10, *n* = 84558 Q(df = 18) = 31.704,

	p = .066	p = .230	p = .024
Fertility: All samples	r = .003 (-0.253, 0.260), p = .980, q = .543 k = 3, s = 3, n = 437 Q(df = 2) = 5.416, p = .067	**r = .130 (0.060, 0.201)**, **p < .001, q = .002** k = 8, s = 6, n = 813 Q(df = 7) = 4.840, p = .679	r = .032 (-0.065, 0.130), p = .514, q = .406 k = 13, s = 5, n = 84128 Q(df = 12) = 17.757, p = .123
RS: All samples	s = 2	r = .192 (-0.052, 0.441), p = .122, q = .189 k = 6, s = 4, n = 205 Q(df = 5) = 11.344, p = .045	**r = .174 (0.085, 0.267)**, **p < .001, q = .002** k = 6, s = 5, n = 430 Q(df = 5) = 0.976, p = .965
Repr. domain: Low fert. samples	s = 0	s = 1	r = .083 (-0.023, 0.190), p = .126, q = .191 k = 8, s = 4, n = 84034 Q(df = 7) = 13.988, p = .051
Fertility: Low fert. samples	s = 0	s = 1	r = .052 (-0.065, 0.169), p = .386, q = .369 k = 7, s = 3, n = 83845 Q(df = 6) = 8.335, p = .215
RS: Low fert. samples	s = 0	s = 0	s = 1
Repr. domain: High fert. samples	r = .099 (-0.012, 0.211), p = .081, q = .140 k = 5, s = 5, n = 1411 Q(df = 4) = 8.799, p = .066	**r = .163 (0.104, 0.225)**, **p < .001, q = .001** k = 13, s = 7, n = 626 Q(df = 12) = 12.347, p = .418	r = .083 (-0.039, 0.205), p = .184, q = .257 k = 11, s = 6, n = 524 Q(df = 10) = 12.595, p = .247
Fertility: High fert. samples	r = .003 (-0.253, 0.260), p = .980, q = .543 k = 3, s = 3, n = 437 Q(df = 2) = 5.416, p = .067	**r = .165 (0.095, 0.237)**, **p < .001, q = .001** k = 7, s = 5, n = 542 Q(df = 6) = 0.988, p = .986	s = 2
RS: High fert. samples	s = 2	r = .192 (-0.052, 0.441), p = .122, q = .189 k = 6, s = 4, n = 205 Q(df = 5) = 11.344, p = .045	**r = .170 (0.053, 0.291)**, **p = .005, q = .022** k = 5, s = 4, n = 241 Q(df = 4) = 0.965, p = .915

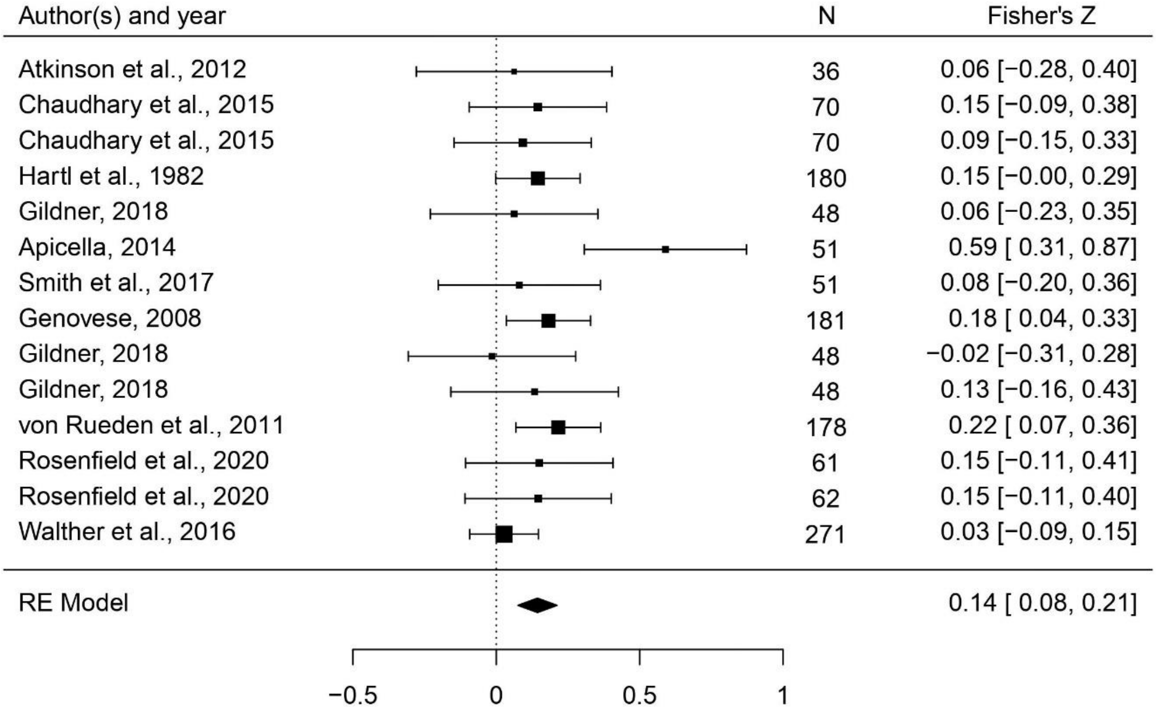
	Voice pitch	Height	T levels
Repr. domain: All samples	r = .136 (-0.053, 0.328), p = .158, q = .228 k = 5, s = 3, n = 143 Q(df = 4) = 5.378, p = .251	r = .006 (-0.049, 0.062), p = .819, q = .491 k = 35, s = 25, n = 22326 Q(df = 34) = 433.359, p < .001	r = .039 (-0.067, 0.145), p = .474, q = .385 k = 3, s = 3, n = 351 Q(df = 2) = 0.387, p = .824
Fertility: All samples	s = 2	r = .011 (-0.039, 0.062), p = .660, q = .451 k = 26, s = 23, n = 22242 Q(df = 25) = 400.038, p < .001	s = 2
RS: All samples	s = 2	r = -0.044 (-0.201, 0.113) , p = .584, q = .430 k = 9, s = 9, n = 603 Q(df = 8) = 33.311, p < .001	s = 1
Repr. domain: Low fert. samples	s = 0	r = -0.037 (-0.112, 0.038), p = .337, q = .347 k = 8, s = 8, n = 17135 Q(df = 7) = 244.970, p < .001	s = 2
Fertility: Low fert. samples	s = 0	r = -0.037 (-0.112, 0.038), p = .337, q = .347 k = 8, s = 8, n = 17135 Q(df = 7) = 244.970, p < .001	s = 2
RS: Low fert. samples	s = 0	s = 0	s = 0
Repr. domain: High fert. samples	r = .136 (-0.053, 0.327), p = .158 , q = .228 k = 5, s = 3, n = 143 Q(df = 4) = 5.378, p = .251	r = .034 (-0.041, 0.109), p = .377, q = .367 k = 27, s = 17, n = 5191 Q(df = 26) = 70.216, p < .001	s = 1
Fertility: High fert. samples	s = 2	r = .059 (0.007, 0.111), p = .025, q = .068 k = 18, s = 15, n = 5107 Q(df = 17) = 26.458, p = .067	s = 0
RS: High fert. samples	s = 2	r = -.044 (-0.201, 0.113), p = .584, q = .430 k = 9, s = 9, n = 603 Q(df = 8) = 33.311, p < .001	s = 1

Note. Fert. = fertility; *k* = number of observations; *n* = number of unique participants; repr = reproductive; RS = reproductive success; Q = Cochran's Q test of heterogeneity; *q* = *q*-value; *s* = number of samples; T = testosterone.

Comparison of high and low fertility samples. The majority (77 %) of observations of reproduction were from high fertility samples. Moderation analyses of low versus high fertility samples could only be conducted for 2D:4D and height; effect sizes did not differ significantly between sample types. Comparing types of high fertility samples (industrialised vs non-industrialised) for 2D:4D and height did not show any differences in effect sizes (see Appendices E). It was not possible to compare sample subtypes for the other traits because observations were almost entirely from non-industrialised populations.

Figure 2.2

Forest plot of the association between body masculinity and the reproductive domain.



Inclusion bias/heterogeneity. Visual inspection of funnel plots (see Appendix H.2) suggested that while the effects for voice pitch, height, and testosterone levels were symmetrically distributed, our analysis may have lacked studies for the other traits. Facial masculinity and height showed significant heterogeneity.

Additional subgroup and moderator analyses for outcome domains. Results of subgroup analyses can be viewed in Table 2.2 and trait-general moderators in Table 2.3; full results of moderation analyses are found in Appendices E.

Moderation analyses (where possible) showed no evidence that the effects of masculinity traits on fertility differed from the effects on reproductive success. However, for body masculinity, the effect on fertility was significant (r = .130, 95% CI: [0.060, 0.201], q = .002; five out of six samples high fertility) while the somewhat larger effect on reproductive success was not. For 2D:4D, there was a significant effect for reproductive success (four out of five samples from high fertility populations: r = .174, 95% CI: [0.085, 0.267], q = .002) but not for fertility.

Similarly, for mating, no trait-general or trait-specific moderators had any consistent effects on the results. Body masculinity effects were stronger where effect sizes had been converted to Pearson's r compared to where they initially had been given as r (B = 0.143, p = .015, q = .047), and effects for height were stronger in gay/mixed sexuality samples than heterosexual samples (B = 0.135, p = .016, q = .047).

2.3.5. Comparing mating and reproduction across traits

Moderation analyses of domain type (mating versus reproduction) for each trait showed no significant differences, although height and testosterone levels had weaker associations with reproduction than mating while body masculinity showed the opposite pattern. There were generally far fewer observations for reproductive measures, so this nonsignificant analysis may reflect lack of power. For facial masculinity, voice pitch, and 2D:4D, effect sizes for global mating and reproductive measures were near identical.

2.4. Discussion

2.4.1. Summary of results

We conducted the first comprehensive meta-analysis of the relationships between men's masculine traits and outcomes related to mating and reproduction. Various proposed (and non-mutually exclusive) hypotheses suggest that more masculine men should show increased mating success (indexed by more matings and/or preferences for short-term mating), increased reproductive output (indexed by fertility and/or reproductive success), and/or lower offspring mortality. Our results showed partial support for these predictions. Global masculinity (i.e., all masculine traits combined) significantly predicted effects in the mating domain, but not the reproductive domain or the offspring mortality domain. When we analyzed each masculine trait separately, all traits except facial masculinity and 2D:4D significantly predicted effects in the mating domain, where similarly strong associations were seen for body masculinity, voice pitch, and testosterone levels, and a weaker correlation was seen for height. In terms of the reproductive domain, the only significant predictor was body masculinity. It was not possible to analyse offspring mortality at the specific predictor level owing to a severe lack of relevant data from which to draw conclusions (total number of observations for each outcome domain: mating domain $k = 371$; reproductive domain $k = 81$; offspring mortality domain $k = 22$).

We also examined how these effects play out in high versus low fertility populations. Typically, however, different outcomes were measured in different groups of populations; mating outcomes were predominantly measured in low fertility populations, while reproductive outcomes were measured mainly in high fertility populations. This made it more challenging to draw direct comparisons. Where it was possible to run moderation analyses on sample type, there were no significant differences. These analyses, however, have small numbers of high and low fertility samples in mating and reproductive outcomes respectively. Therefore, while we can confidently say that most forms of masculinity (but not facial masculinity or 2D:4D) are associated with (largely self-reported) mating outcomes in low fertility samples, we cannot draw any clear conclusions regarding mating success in high fertility samples. Similarly, although we are confident that body masculinity is associated with fertility/reproductive success in high fertility samples, we cannot draw conclusions about low fertility contexts.

More generally, our moderation analyses on outcome types and factors relating to measure quality did not yield any consistent differences between effect sizes, suggesting that the effects we do find are reasonably robust within sample type at least. Two key points to note here are that: *i*. although effect sizes for mating attitudes and mating behaviours did differ for some traits (i.e., facial masculinity and body masculinity), these differences were never significant, despite mating behaviours being constrained by opportunities (assuming participants report truthfully), and *ii*. similarly, effect sizes did sometimes differ by publication status but never significantly so; in addition, the direction of the differences was not consistent (i.e., effect sizes were not consistently larger in published analyses). Even if the analysis was restricted to nonpublished effects only, the association between body masculinity and both mating and reproduction would be weaker but remain significant (mating: $r = .077$, $p = .006$; reproduction: $r = .112$, $p < .001$; both associations would remain significant after q-value computation). Overall, this suggests that researchers have not been selectively reporting larger effect sizes.

Compared to previous meta-analyses assessing associations between handgrip strength and mating outcomes (Van Dongen & Sprengers, 2012), height/strength and reproductive outcomes (von Rueden & Jaeggi, 2016; Xu et al., 2018), and testosterone levels and mating effort (Grebe et al., 2019), our analysis benefits from more comprehensive measures of masculinity, larger sample sizes, and inclusion of more unpublished effects. With the exception of Xu and colleagues' analysis (2018), we observe smaller effect sizes than previous meta-analyses, which suggests that the association between masculinity and fitness outcomes has previously been overestimated. In general, what significant associations we did observe were small and ranged between r = .05 and .17, although they are potentially meaningful in an evolutionary context. As benchmarks for interpreting correlations, Funder and Ozer (2019) suggest that a correlation of .10, while being a small effect, has the potential to be influential over a long time period, and a medium-size correlation of .20 can be consequential both in the short- and long-term. The cumulative effect of relatively 'weak' correlations can therefore be of real consequence, particularly when considered in terms of selection acting over many generations.

2.4.2. Major implications

Selection for body masculinity. The first stand-out result of our analysis is that body masculinity (i.e., strength/muscularity) is the only trait in our analysis that was consistently correlated with both mating and reproductive outcomes across populations, and the effects of body masculinity on these outcomes were among the strongest in the analysis. In contrast, other aspects of masculinity (except facial masculinity and 2D:4D) predicted mating success in low fertility samples but did not yield reproductive benefits in high fertility samples.

Body masculinity is therefore the trait where we have the most compelling evidence that selection is currently happening within naturally fertile populations − and from that, can infer that selection likely took place in prior eras as well. As such, our results are consistent with the argument that dimorphisms in strength and muscle mass are sexually selected. Overall, since traits such as body size, strength, and muscularity are associated with formidability, our findings are consistent with the male-male competition hypothesis. In species with male intrasexual competition, males tend to evolve to become larger, stronger, and more formidable than females, as they are in humans. Some authors argue that male-male violence has influenced human evolution (Hill et al., 2017; Gat, 2015), and male intergroup aggression increases mating/reproductive success in both non-industrialised human societies and in non-human primates (Glowacki & Wrangham, 2015; Manson et al., 1991). (And indeed, the non-human evidence might suggest this form of dimorphism has been under selection since pre-hominid ancestors, although the strength of such selection pressures have likely fluctuated over this time [Plavcan, 2012].) For example, in the Yanomamö Indians, men who kill others have greater reproductive success (Chagnon, 1988). A relationship between formidable traits and fitness outcomes need not be a direct one, however. It might, as mentioned in the introduction, be mediated by other factors that are important in mate choice, such as interpersonal status and dominance. For example, features that are advantageous in intraspecies conflicts may also be advantageous when hunting game (Sell et al., 2012); Smith and colleagues (2017) reported that in a hunter-gatherer population, men with greater upper body strength and a low voice pitch had increased reproductive success, but this relationship was explained by hunting reputation.

It is of course possible that different selection pressures may have contributed to the evolution of different masculine traits. Male-male competition for resources and mates, female choice, and intergroup violence are all plausible, non-mutually exclusive explanations (Plavcan, 2012). In this article, we have focused on the effect of men's own traits on their fitness, but it is of course equally possible that men varying in masculinity may differ in the quality of the mates they acquire. If masculine men are able to secure mates who are more fertile and/or better parents, this may also increase their fitness.

No evidence of advantage for facial masculinity. Considerable attention has been given in the literature to the hypothesis that masculinity in men's facial structure is an indicator of heritable immunocompetence (i.e., good genes), which should then be associated with greater mating and reproductive success. While we find that the effect of facial masculinity on mating was similar in size to that of other traits (*r* = .08), it was not significantly different from zero, suggesting more variability in effects. Furthermore, the effect of facial masculinity on mating (such as it was) was largely driven by mating attitudes and was close to zero for mating behaviours, suggesting that men's facial masculinity exerts virtually no influence on mating when moderated by female choice. Similarly, the influence of facial masculinity on fertility in high fertility samples was non-existent (*r* = .00). Although the relationship with reproductive success appeared stronger, this was based on only two samples. This is, all together, doubly striking because although voice pitch, height, and testosterone levels did not predict reproductive outcomes, they did all relate to mating in the expected direction. Facial masculinity is ergo an outlier in being so entirely unrelated to mating success in our data, while subject to so large a literature assuming the opposite.

Overall, these findings contradict a large body of literature claiming that women's preferences for masculinity in men's faces are adaptive. Rather, they indicate that such preferences (to the extent they exist at all) are a modern anomaly only found in industrialised populations, as suggested by Scott and colleagues (2014), and as demonstrated by the positive correlation between facial masculinity preferences and national health and human development indices (Marcinkowska et al., 2019).

Students and foragers. One key observation regarding our dataset is that it shows a rather 'bimodal' distribution between a large number of studies sampling (predominantly English-speaking) students on one hand, and a cluster of studies sampling foragers, horticulturalists, and other subsistence farmers (predominantly from just two continents) on the other. Where it was possible to compare student vs non-student/mixed samples within low fertility populations, and traditional vs industrialised high fertility samples, we generally did not find any differences. Likewise, where it was possible to compare monogamous and formally polygynous cultures, we also found no differences. This is despite evidence that monogamy actually changes selection pressures on human men (Brown et al., 2009). Therefore, although we are reasonably confident that our results regarding body masculinity and reproduction are robust, insofar as they are based on non-industrialised populations with a range of subsistence patterns (hunter-gatherers, forager-horticulturalists, and pastoralists), it remains essential to consider rebalancing the literature. Not only do we require more holistic representation of non-industrialised populations (drawing from Asia and Oceania in particular, where we had one and zero samples, respectively), but it is also important to increase representation of non-student participants in low fertility contexts.

Disconnection between mating and reproductive literatures. As noted above, we found that voice pitch, height, and testosterone levels were associated with (largely self-reported) mating success in mostly low fertility populations, but not with actual reproductive fitness in high fertility populations. A caveat here is that effect sizes for voice pitch and reproduction were similar in strength to effect sizes for body masculinity, but we note that this analysis had the smallest sample size of our whole analysis ($k = 5$, $N = 143$), which prevents us from drawing firm conclusions regarding the relationship between voice pitch and reproductive outcomes.

Overall, however, the contradicting pattern of results for the traits mentioned above raise important concerns for the human sexual selection field, particularly with respect to whether (and which) mating measures can be used as reliable indicators of likely ancestral fitness when considering the current evidence base. Since reproductive outcomes – for good reason – are not considered meaningful fitness measures in populations with widespread contraception use, we typically test fitness outcomes in industrialised populations using mating measures such as sociosexual attitudes and casual sexual encounters. This is done under the assumption that such measures index mating strategies that ancestrally would have increased men’s offspring numbers. However, if mating outcomes (be it attitudinal or behavioural) measured in low fertility populations truly index reproductive outcomes in naturally fertile contexts, we would expect traits that predict mating to also predict reproduction on average across samples (notwithstanding the diversity in norms/reproductive behaviours across high fertility samples). We do not, however, have evidence that this is generally the case. Our findings therefore raise the question of whether these widely used measurements are truly valid proxies of what we purport to be measuring.

Our findings thus illustrate that when we attempt to test the same underlying research questions using different measurements in different populations, this may yield conclusions that are erroneous or misleading when applied outside of the studied population. We suggest, based on our analysis, that researchers could for instance consistently gather sexual partner number, age of marriage, and number/survival rates of offspring in multiple population types. Wherever possible, it is essential to use the same measurements across populations, or at least resist the temptation of applying our findings universally.

2.4.3. Key limitations

Non-linearity. A limitation of our analysis is that we only assessed linear relationships, ignoring possible curvilinear associations. There is evidence suggesting that moderate levels of masculinity might be associated with increased reproductive success (see e.g., Boothroyd et al., 2017, for offspring survival rates) and perceived attractiveness (Frederick & Haselton, 2007; Johnston et al., 2001; but see also Sell et al., 2017), with a decrease for both very low and very high levels of masculinity. Indeed, some of these authors have argued that masculinity may be under stabilising, rather than directional, selection in humans. In instances such as these, our ‘null’ conclusions regarding e.g., facial masculinity, remain valid; facial masculinity does not appear to be under directional selection. However, we also note that there is data suggesting that height in men may be optimal when it is over-average but not maximal. In this scenario, although the linear relationship would be weaker, the trait remains under directional selection, and we would still expect to see positive, albeit weak, associations in our analyses. In the vast majority of studies included, only linear relationships were tested, and acquiring original data to investigate and synthesise non-linear effects was beyond the scope of the current article. However, increased publication of open data with articles may well facilitate such a project in future years.

Testosterone effects. As mentioned above, in our analysis testosterone levels predicted mating outcomes – with similar effect sizes for attitudinal and behavioural measures – but did not predict reproduction. While a causal relationship between testosterone levels and mating success cannot be established from this (i.e., whether high testosterone men pursue more mating opportunities which leads to more matings, or whether high testosterone results from many matings), testosterone is commonly argued to motivate investment in mating effort. *If* current testosterone levels index degree of masculine trait expression in men, our results *might* indicate that masculine men's increased mating success is due to greater pursuit of matings − rather than reflecting female choice and/or greater competitiveness. Two caveats for interpreting our results, however (applicable both to the significant effect we observe for mating and the nonsignificant effect for reproduction), is that circulating testosterone levels *i.* change over the course of a man's lifetime, peaking in early adulthood and subsequently declining (Booth & Dabbs Jr, 1993; although this may not be the case in non-industrialised populations: Bribiescas, 1996), and *ii.* are reactive. In the studies we gathered, testosterone levels were generally measured contemporaneously with mating/reproductive data collection – not when masculine traits generally become exaggerated in adolescence. Testosterone also decreases, for example, when men enter a relationship or get married (Archer, 2006; Holmboe et al., 2017), when they become fathers (Archer, 2006; Gettler et al., 2011), or when they engage in childcare (Archer, 2006). Thus, men whose testosterone levels were previously high may show declining testosterone levels either because of their age and/or because their relationship or fatherhood status has changed. This limits the conclusions we can draw, both with regards to a potential mediating role of testosterone levels in the association between masculine traits and mating success, and the observed non-existent effect for testosterone levels and reproductive outcomes. We also note that the sample size for reproduction, as a function of testosterone levels, was small.

2.4.4. Conclusion

In summary, we used a large-scale meta-analysis of six masculine traits and their relationships with mating and reproductive outcomes to test whether such traits are currently under selection in humans. We found that all masculine traits except facial masculinity and 2D:4D were associated with significantly greater mating success. However, only body masculinity predicted higher fertility, indexed by reproductive onset, number of offspring, and grand-offspring. We further note that the mating and reproduction literature is starkly split between studying mating in predominantly student settings, and ‘only’ fertility in high fertility settings, which imposes constraints on both this paper and our field as a whole. We argue that our findings illustrate that when we test hypotheses about human evolution largely in industrialised populations, we risk drawing conclusions that are not supported outside of evolutionarily novel, highly niche mating and reproductive contexts. We therefore call for greater sample diversity and more homogenous measurements in future research.

Chapter 3.

Maternal morphology and parental attractiveness as indicators of reproductive outcomes

---

3.1. Introduction

3.1.1. Sexually dimorphic traits in women

One potential pathway through which men can increase their biological fitness is by mating with 'better quality' partners, where quality is typically indexed in research by sexually dimorphic/attractive traits. Dimorphic, i.e., feminine, traits in women include facial morphology and body shape/fat distribution. Compared to men, women have on average a more neotenous facial structure, with a more oval face shape, a longer forehead, larger eyes, a smaller nose, a smaller chin, and fuller lips (Perrett et al., 1994; Rhodes, 2006). Both the amount and distribution of body fat differs between men and women: while men in most populations typically have an average body fat percentage around 10-15%, women display an average fat percentage twice as high, ranging between 20 and 30% (Wells, 2007). Women also store fat primarily in the gluteofemoral region, i.e., on the hips, buttocks, and thighs, while men accumulate fat mainly in the abdominal area (Norgan, 1997). This dimorphism in body fat distribution also contributes to women of childbearing age displaying a more hourglass-shaped body with a lower waist-to-hip ratio (WHR), with ratios in white women typically falling between .67-.80 compared to between .85-.95 in men (e.g., Marti et al., 1999). Adult women also have permanently enlarged breasts (Marlowe, 1998), in addition to the sex differences in height, musculature and strength, facial and body hair growth, voice pitch, and 2D:4D discussed in previous chapters.

Women's feminine traits generally develop under the influence of oestrogen. For example, whereas male and female WHRs are very similar in childhood, girls' increasing pubertal oestrogen levels stimulate fat deposits in the gluteofemoral area, resulting in them developing a more hourglass-shaped body (reviewed in Wells, 2007). These fat deposits are subsequently used during lactation (Norgan, 1997). Additionally, evidence suggests that gluteofemoral fat deposits store the omega-3 fatty acid docosahexaenoic acid, which has been suggested by some to be particularly important for infant brain development (reviewed in e.g., Lassek & Gaulin, 2008; 2019). Pubertal increases in oestrogen also cause the breasts to grow, while simultaneously hindering skeletal growth (Bordini & Rosenfield, 2011). The latter results in women having a shorter stature (reviewed in Dunsworth, 2020) and also leaves them with a more juvenile facial structure which undergoes fewer pubertal changes compared to the changes seen in men's faces (Rhodes et al., 2003). Such feminine traits, particularly facial femininity and a WHR in the lower end of the normal range, are typically found to be attractive (e.g., Brooks et al., 2015; Dixson et al., 2010; Fiala et al., 2021; Perrett et al., 1994; Rhodes, 2006; Thornborrow et al., 2018; but see also Scott et al., 2014, and Boothroyd et al., 2021).

3.1.2. The association between women's oestrogen levels/morphological traits and health/immunity

While the ICHH is a widely cited (yet also criticised) explanation of the evolution of masculine traits in men, there exists at present no similarly prevalent explanation of feminine traits cueing immunocompetence in women. Oestrogen does play a role in women's immunity, however. Immune function is sexually dimorphic (Roved et al., 2017), with women generally showing lower infection rates and lower overall mortality but higher rates of autoimmune illnesses (Olsen & Kovacs, 1996). This is likely due to general immunoenhancing effects of oestrogen, hyperactivating the immune system (Zandman-

Goddard et al., 2007). In contrast, oestradiol (a type of oestrogen) is commonly claimed to have immunosuppressive properties (e.g., Hodges-Simeon et al., 2019; Jansson & Holmdahl, 1998; but see also van Anders, 2010, who reported a positive link between women's oestradiol levels and immune response). Thus, while oestrogen may enhance and suppress different aspects of immunity, respectively, women generally have better immunity than men and the evidence for oestrogen's immunoenhancing role is stronger compared to testosterone (Foo et al., 2017a; Hodges-Simeon et al., 2019).

Does this mean that oestrogen-dependent, i.e., feminine, traits cue better immune functioning? In women, facial femininity is strongly correlated with attractiveness (e.g., Fiala et al., 2021; Rhodes et al., 2003). Neither femininity nor attractiveness have shown consistent links with health outcomes, however (see Table 3.1). Rated facial femininity has been reported to be unrelated to childhood and adolescent health (Rhodes et al., 2003), while it has also been found that adult morphometric femininity was positively associated with a limited number of immunity measures at age 13, largely reflecting improved allergic responses and antibacterial immunity (Foo et al., 2020). Kalick and colleagues (1998) observed no relationship between late adolescent facial attractiveness and health measured at various timepoints in adolescence and adulthood. It is possible, however, that rather than *increased* attractiveness signalling *good* health, being *below* median attractiveness might signal *poor* health (Zebrowitz & Rhodes, 2004). Furthermore, Henderson and Anglin (2003) found that facial attractiveness assessed at age 17 positively predicted longevity. In two studies measuring self-reported common illnesses, more facially feminine women reported fewer respiratory infections, no difference in terms of gastrointestinal infections, and fewer versus no relationship with antibiotic use (Gray & Boothroyd, 2012; Thornhill & Gangestad, 2006). In one of the latter studies, attractiveness was correlated with less antibiotic use and marginally fewer respiratory infections, but showed no relationship with health in the other study. Elsewhere, attractive women have reported lower prevalence of a composite of common health problems (Hume & Montgomerie, 2001), whereas Shackelford and Larsen (1999) observed that attractive women reported significantly fewer headaches but no difference in terms of other common illnesses. Foo and colleagues (2017b) found that neither rated femininity nor attractiveness predicted bacterial killing capacity, overall bacterial immunity, lysozyme activity or bacterial suppression capacity; Cai and colleagues (2019) similarly reported no relationships between morphometric masculinity (with facial masculinity corresponding to the opposite of femininity) or attractiveness with common illnesses or salivary immunoglobulin A. Furthermore, no associations with facial attractiveness/femininity have been reported for women's post-vaccination antibody response (Rantala et al., 2013b) nor for heterozygosity in the major histocompatibility locus (MHC: Coetzee et al., 2007; Lie et al., 2008). Lastly, Żelaźniewicz and colleagues (2020) observed positive associations between facial attractiveness and lipid but not glucose homeostasis, liver functioning or inflammation markers. To date, the evidence for any consistent relationship between women's facial traits and immunity measures is therefore scant.

Table 3.1

Summary of associations between female dimorphism/attractiveness and health/immunity outcomes.

Facial traits					
Author(s) & year	Sample	Trait	Outcome	Notes	Effect
Kalick et al., 1998	Inter-generational Studies, U.S. (born 1920-29)	Attr.	Childhood/ adolescent health	Attr. rated at age 17-18 yrs. Health rated by physicians through clinical exams and histories at 11-18 years	Null
			Middle adulthood health	Health based on a medical exam and history at 30-36 years	Null
			Later adulthood health	Heath rated by physicians at 58-66 years	Null
Shackelford & Larsen, 1999	University students, U.S.	Attr.	Headache	Self-reported	-ve (one-tailed): $r = -.26$

			Runny or stuffy nose		Null
			Nausea or upset stomach		Null
			Muscle soreness or cramps		Null
			Sore throat or cough		Null
			Backache		Null
			Jitteriness or trembling		Null
			Cardiovascular health	Measured by cardiac recovery time	Null
Hume & Montgomerie, 2001	University students, Canada	Attr.	Common health problems	Health assessed by composite of fractures, various diseases, and cosmetic or reconstructive surgeries	-ve: β = -.30
Rhodes et al., 2003	Same as Kalick et al., 1998	Rated fem.	Childhood/ adolescent health	Health rated by physicians through clinical exams and histories at 11-18 years	Null
Zebrowitz & Rhodes, 2004 (re-analysis of Kalick et al., 1998)	Same as Kalick et al., 1998	Attr. above median	Childhood/ adolescent health	Attr. assessed in adolescence (men and women analysed together throughout)	Null
		Attr. below median	Childhood/ adolescent health	Attr. assessed in adolescence	Null
		Attr. above median	Childhood/ adolescent health	Attr. assessed in adulthood	Null
		Attr. below median	Childhood/ adolescent health	Attr. assessed in adulthood	+ve: *r* = .30
Henderson & Anglin, 2003	High school students, Canada (born ~1907-10)	Attr.	Longevity	Attr. assessed through photographs from high school yearbooks, ~17 years	+ve: *r* = .36
Thornhill & Gangestad, 2006	University students, U.S.	Measured masc.	Number of respiratory infections	Illnesses measured in the past 3 years, self-reported	+ve: *r* = .18
			Days of respiratory infections		+ve: *r* = .19
			Number of gastrointestinal infections		Null
			Days of gastrointestinal infections		Null
			Antibiotic use		Null
		Attr.	Number of respiratory infections	Illnesses measured in the past 3 years, self-reported	Null
			Days of respiratory infections		Null
			Number of gastrointestinal infections		Null
			Days of gastrointestinal infections		Null
			Antibiotic use		Null
Coetzee et al., 2007	University students, South Africa	Attr.	MHC heterozygosity		Null

Lie et al., 2008	University students, Australia	Rated fem.	Mean MHC heterozygosity		Null
			Standardised mean genetic distance between alleles		Null
		Attr.	Mean MHC heterozygosity		Null
			Standardised mean genetic distance between alleles		Null
Gray & Boothroyd, 2012	University students, U.K.	Rated fem.	Bouts of respiratory infections	Illnesses measured in the past 3 years, self-reported	Null
			Days of gastroenteritis		Null
			Days of antibiotic use		-ve: $r = -.272$
			Bouts of respiratory infections	Illnesses measured in the past year, self-reported	-ve: $r = -.215$
			Days of gastroenteritis		Null
			Days of antibiotic use		-ve: $r = -.211$
			Bouts of colds	Illnesses measured at follow-up, self-reported	Null
			Days of flu		-ve: $r = -.265$
			Days of gastroenteritis		Null
			Days of antibiotic use		Null
			Days off for illness		Null
		Attr.	Bouts of respiratory infections	Illnesses measured in the past 3 years, self-reported	Null
			Days of gastroenteritis		Null
			Days of antibiotic use		-ve: $r = -.226$
			Bouts of respiratory infections	Illnesses measured in the past year, self-reported	Null
			Days of gastroenteritis		Null
			Days of antibiotic use		Null
			Bouts of colds	Illnesses measured at follow-up, self-reported	Null
			Days of flu		Null
			Days of gastroenteritis		Null
			Days of antibiotic use		Null
			Days off for illness		Null
Rantala et al., 2013b	Latvian sample	Attr.	Hepatitis B antibody response		Null
Foo et al., 2017b	University students, Australia	Rated fem.	Bacterial killing capacity/ overall bacterial immunity/ lysozyme activity composite		Null
			Bacterial suppression capacity		Null
		Attr.	Bacterial killing capacity/ overall bacterial immunity/lysozyme activity		Null
			Bacterial suppression capacity		Null
Cai et al., 2019	University students, U.K.	Morph. masc.	Infection frequency and recency	Self-reported infections in the past year	Null

			Upper respiratory illnesses		Null
			Perceived vulnerability to disease		Null
			Salivary immuno-globulin A		Null
		Attr.	Infection frequency and recency		Null
			Upper respiratory illnesses		Null
			Perceived vulnerability to disease		Null
			Salivary immuno-globulin A		Null
Foo et al., 2020	The Western Australian Pregnancy Cohort (Raine) Study	Morph. fem.	Immunity composite 13 (largely allergic responses)	Femininity measured at 22 years, immunity at 14 years. Immunity measures across 16 composites included haematology, antibodies, cytokine responses, inflammatory markers	+ve: $r = .19$
			Immunity composite 15 (antibacterial immunity)		+ve: $r = .17$
			14 separate immunity composites		Null
Żelaźniewicz et al., 2020	Urban Western sample	Attr.	Lipid homeostasis		-ve: $\beta = - .17$
			Glucose homeostasis		Null
			Liver functioning		Null
			Inflammation markers		Null
			Bodily traits		
Pawlowski et al., 2014	University students, Poland	WHR	Pathogen colonisation		-ve: $H = 54.9$
Ellis & Hoskin, 2020	University students, Malaysia	Strength	Self-reported health		+ve: $\rho = .520$
		Musc-ularity	Self-reported health		+ve: $\rho = .103$
		2D:4D (inverted)	Self-reported health		Null
	University students, U.S.	Strength	Self-reported health		+ve: $\rho = .331$
		Musc-ularity	Self-reported health		+ve: $\rho = .242$
		2D:4D (inverted)	Self-reported health		Null

Note. MHC = major histocompatibility complex. Since the predicted direction of effects differ between studies depending on measures used, effects showing significantly better health as a function of *increased femininity* are indicated in green and worse health in orange. We do not indicate marginally significant effects as significant here (as is sometimes done in the papers where they were reported). Note that we have not included measures of height here, nor of cardiovascular health as a function of WHR.

Women's body traits have rarely been studied in relation to *immunity* per se. The majority of research in this area has instead focussed on other health outcomes, typically as a function of women's WHR (reviewed in Bovet, 2019). Women with a higher, i.e., less feminine, WHR are commonly claimed to show poorer cardiovascular (reviewed in Björntorp, 1987; but see also Lassek & Gaulin, 2018) and reproductive health (reviewed in Dağ & Dilbaz, 2015). It should be noted, however, that the poor health outcomes associated with high WHRs are usually seen in obese individuals. Since obesity is unlikely to have existed pre-agriculture (Eaton et al., 1997) and thus is an evolutionarily novel phenomenon, it is implausible that men's preferences for low WHRs in women evolved because abnormally high WHRs cue poor health (Lassek & Gaulin, 2018). The only study, to our knowledge, to assess WHR in relation to immunity found that women with a more feminine WHR showed *poorer* immunity (Pawlowski et al., 2014).

Aside from WHR, Ellis and Hoskin (2020) reported that women's self-reported muscularity, strength, and height (i.e., traits that are pronounced in men, not women), but not 2D:4D, significantly predicted better self-reported health. Generally, however, women's height does not appear to be linked to better immunity (Foo et al., 2020; Krams et al., 2014; Pawlowski et al., 2014) even if height is related to other health variables in both men and women (e.g., Stefan et al., 2016).

3.1.3. The association between women's oestrogen levels/morphological traits and reproductive outcomes

Oestrogens, such as oestradiol, influence fertility (Baird et al., 1997; Lipson & Ellison, 1996; Lubahn et al., 1993; Sher & Rahman, 2000; Venners et al., 2006). Due to their putative association with not only pubertal but also adult levels of oestrogen, feminine traits are commonly posited to cue reproductive health (reviewed in Weeden & Sabini, 2005); in short, if feminine traits cue current hormone levels, they should also cue fertility. Under this hypothesis, men should gain reproductive benefits by mating with feminine women, not so much in terms of siring better-quality offspring, but primarily by increasing their own reproductive output. In spite of this claim attracting considerable research attention, evidence for a link between adult hormone levels and femininity is mixed. Positive relationships with oestrogen levels have been reported between facial femininity (Law Smith et al., 2006), facial (Durante & Li, 2009; Law Smith et al., 2006, Żelaźniewicz et al., 2020) and bodily attractiveness (Durante & Li, 2009; Grillot et al., 2014), body shape (Jasieńska et al., 2004) and right-hand but not left-hand 2D:4D (McIntyre et al., 2007). Other authors, however, have reported no associations between women's ovarian hormone levels and body shape (Grillot et al., 2014; Jones et al., 2018; Rilling et al., 2009) or attractiveness (Jones et al., 2018; Rilling et al., 2009). Furthermore, associations with testosterone have sometimes been observed: facially attractive women may have lower testosterone levels overall or a low testosterone-to-estradiol ratio (Probst et al., 2016; Żelaźniewicz et al., 2020). In another study, however, women with the lowest WHRs had high levels of both testosterone and estradiol, but this was only the case for women in the fertile phase of the menstrual cycle (Mondragón-Ceballos et al., 2015). Similarly, van Anders and Hampton (2005) reported a weak but significant positive relationship between women's WHR and testosterone levels. Null relationships between testosterone levels and body shape/body attractiveness have also been reported, however (Grillot et al., 2014).

Crucially, the hypothesised relationship between feminine traits and reproductive outcomes requires empirical evidence. Despite the prevalence of this claim, it is a severely understudied association. Furthermore, in order to establish the presence of such an association, longitudinal designs would be required where morphological traits are measured prior to reproduction, and fertility data is collected longitudinally. This is particularly pertinent to determine an association between women's body shape and fertility, as women's bodies – and particularly their WHR – change post-reproduction (Wells et al., 2010). Doubtless for practical reasons, however, traits and fertility are usually measured simultaneously (e.g., Butovskaya et al., 2017), which limits the conclusions that can be drawn from such studies.

3.1.4. The present chapter

The hypotheses outlined above predict that women's feminine traits should be associated with better reproductive outcomes: *i.* if femininity cues better immunity/health, then feminine women should produce healthier, more viable offspring, and *ii.* if femininity cues fertility, then feminine women should produce a greater quantity of offspring. The latter could be achieved not only by being more fecund, but alternatively – or additionally – by starting to reproduce at a younger age. To reiterate, neither of these predictions have been sufficiently tested and researchers typically rely on using proxies of fertility rather than measuring actual reproduction.

In the present chapter, we firstly report the results of two studies (Studies 2 and 3) where we analysed the associations between morphological traits and reproductive outcomes in two archival non-Western datasets, hypothesising that increased parental dimorphism and attractiveness will predict greater offspring numbers and offspring survival. In Study 2, we report associations between facial traits (femininity, attractiveness, and perceived health) and reproductive outcomes, indexed by reproductive onset, number of births, and offspring survival, in a sample of Agta women. In Study 3, we analysed the associations between parental facial traits (masculinity/femininity, attractiveness, and perceived health) and offspring numbers in a sample of Turkish parents. Next, we conducted a systematic literature review and meta-analysis of reproductive outcomes as a function of morphological and attractive traits in women and attractive traits in men, including our data from Studies 2 and 3. In Chapter 2 (Study 1), our meta-analysis of men's masculine traits showed that men's physical strength and muscularity positively predicted reproductive outcomes. To rule out the possibility that such a relationship is mediated by maternal strength/muscularity, in this review we also included traits that are typically exaggerated in men, such as strength, in addition to traits that are exaggerated in women.

Both the ICHH (applied to men) and the hypotheses outlined above (applied to women) assume that dimorphic traits should be perceived as attractive in both sexes, but we did not include men's attractiveness in our meta-analysis. Here, we therefore included attractiveness as a hypothesised predictor of reproductive outcomes in both men and women (Study 4). Outcomes thus include both fertility and offspring viability; the analysis in women thereby mirrors our meta-analysis in men, which is in turn extended by including attractiveness as a predictor of reproduction in both sexes.

3.2. Study 2: Agta women

3.2.1. Methods

3.2.1.1. Participants

Participants were women from the Agta Demographic Database (Headland et al., 2011). The database comprises data collected from 4,300 individuals in the Agta population, which has been studied over many years by Thomas and Janet Headland (reported in e.g., Headland, 1986; 1989). The population live on the Luzon Island of the Philippines and were traditionally hunter-gatherers, residing in seminomadic small camps, but in the 1970s transitioned more to a peasant economy. They are monogamous, with high fertility and mortality: the total fertility rate is 7.0 and life expectancy at birth is only 23 years (Headland et al., 2011). Here, we report data from 119 women for whom ratings of facial photographs and fertility data exist. Men in the database have previously been reported by Boothroyd and colleagues (2017), who found that offspring of men with intermediate levels of facial masculinity were most likely to survive, but no other significant associations between men's facial traits and reproductive outcomes.

3.2.1.2. Materials and procedure

The procedure for the collection of facial ratings was approved by Durham University Department of Psychology Ethics Committee.

Women's facial images. Photographs were extracted from the databases. They were primarily black-and-white and front-facing, but head tilt and angle varied between photographs. Facial expressions were typically neutral, but this did vary. Photographs were taken outside, and lighting conditions therefore differed between photographs. Participants' age at the time of being photographed ranged between 17-64 years of age (M = 34.61, SD = 12.24).

Reproductive variables. We extracted the following variables from the database: age at first pregnancy, age at first birth, number of liveborn offspring, total number of births (including stillbirths), and the proportion of offspring surviving until 5 years of age. For age at first pregnancy and age at first birth, we excluded cases where age was set to < 12 years (omitting one case) as this was deemed erroneous, leaving a final sample of 118 women; 16 of these were of reproductively active age when their reproductive data were last updated. All reproductive variables were skewed; for age at first pregnancy and age at first birth, log 10-transformation reduced but did not eliminate skew.

Rated facial traits. Images had previously been rated by observers for femininity, attractiveness, and perceived health (reported in Boothroyd et al., 2017). In total, 13 British observers (6 male; age range of all observers 16-28 years) rated the images. Images were masked to only show the outline of the face, excluding the hair, neck, and ears. Image order was randomised, and observers rated the images in the lab or online. Average ratings were calculated for each face on each trait. All facial trait scores were normally distributed.

Table 3.2

Descriptive statistics for all study variables in the Agta women (Study 2).

Mother's traits	*N*	Min	Max	Mean	*SD*
Facial femininity	118	2.46	6.08	4.05	0.80
Facial attractiveness	118	1.69	5.46	3.09	0.70
Facial health	118	2.00	5.69	3.73	0.79
Reproductive variables	***N***	**Min**	**Max**	**Mean**	***SD***
Age at first pregnancy	95	14	42	20.85	4.52
Age at first birth	94	14	42	20.99	4.57
Number of liveborn offspring	104	1	12	5.50	3.01
Total number of births	105	1	12	5.75	3.16
Proportion of offspring surviving until age 5	95	0.00	1.00	0.71	0.26

3.2.1.3. Statistical analysis

Means and standard deviations for all study variables are shown in Table 3.2. Age at the time of being photographed was negatively correlated with all facial traits (femininity: $r = -.366$, $p < .001$; attractiveness: $r = -.458$, $p < .001$; perceived health: $r = -.510$, $p < .001$). Thus, we saved the residuals from separate linear regressions with age in photograph as the predictor variable and rated facial femininity, attractiveness, and perceived health, respectively, as the outcome variable. We then ran Pearson's correlations between the saved residuals and the reproductive variables. Current age was not significantly associated with reproductive outcomes and was therefore not controlled for. Lastly, we checked for quadratic associations by fitting curve estimations.

3.2.2. Results

All facial traits, corrected for age, were significantly and positively correlated with each other: femininity and attractiveness $r = .625$, $p < .001$, $q = .011$; femininity and perceived health $r = .597$, $p < .001$, $q = .011$; attractiveness and perceived health $r = .838$, $p < .001$, $q = .011$.

Overall, the results did not support the predictions (Table 3.3). Contrary to expectations, no facial traits were associated with younger age at first pregnancy or first birth and effect sizes were close to zero for these associations. Furthermore, while all facial traits showed negative correlations with number of births (both liveborn and total), the only significant association of these was between perceived facial health and number of liveborn births ($r = -.208$, $p = .034$, $q = .310$). In line with the hypothesis, all facial traits correlated positively with proportion of offspring who survived until age 5; the only significant predictor was facial attractiveness ($r = .214$, $p = .037$, $q = .310$), however. Neither of the significant relationships survived corrections for multiple comparisons. We detected no significant quadratic associations.

Table 3.3

Pearson's bivariate correlations between mother's facial traits and reproductive variables in the Agta women, controlling for age in photograph (Study 2).

Mother's traits	Age at first pregnancy	Age at first birth	No of liveborn offspring	Total no of births	Prop. of offspring surviving
Facial femininity	-.023 *N* = 95	.005 *N* = 94	-.157 *N* = 104	-.113 *N* = 105	.083 *N* = 95
Facial attractiveness	-.039 *N* = 95	.011 *N* = 94	-.122 *N* = 104	-.078 *N* = 105	.214* *N* = 95
Facial health	.009 *N* = 95	.054 *N* = 94	-.208* *N* = 104	-.147 *N* = 105	.201 *N* = 95

Note. Prop = proportion.
* $p < .05$. No associations remained significant after computation of *q*-values.

3.2.3. Interim discussion

Women's feminine traits, such as facial femininity and attractiveness, are often claimed to cue higher fertility. Here, we observed that women's facial femininity, attractiveness, and perceived health did not predict greater fertility indexed by earlier reproductive onset in a sample of women from the small-scale Agta population. More facially healthy-looking women had given birth to fewer liveborn offspring, however, and more facially attractive women had a higher proportion of offspring surviving until age 5 – but neither relationship remained after correcting for multiple comparisons.

It is noteworthy that all facial traits showed negative correlations with number of births (both liveborn and total births), indicating the possibility that in a larger sample, women who look more facially feminine, attractive, and healthy may have *lower* rather than higher fertility. If this is the case, this stands in contrast to the common claim that women's feminine and attractive traits index higher fertility. Since these associations are nonsignificant, however, they should be interpreted with caution. While that is also the case for the positive relationship between maternal facial attractiveness and offspring survival – which did not survive corrections – *if* it is an association which would remain significant in a higher-powered sample, it *could* suggest that the offspring of more attractive women are of better quality themselves and therefore more viable. Alternatively, it could suggest the presence of other mediating variables, such as quality of parental care. Whereas we found no linear relationship between men's masculinity and offspring mortality in Chapter 2, Boothroyd and colleagues (2017) reported that the lowest offspring mortality was found for fathers with intermediate levels of facial masculinity in the men in the same population. This might suggest a mediating role of paternal quality – but showing a quadratic rather than the typically hypothesised linear relationship between paternal trait expression and offspring quality.

Lastly, it is also possible that, if more feminine and attractive women do have fewer offspring overall (ignoring a potential causal association between number of births and attractive facial appearance), that the offspring these women do have benefit from less competition from siblings, and therefore are more likely to survive – particularly in a pre-industrial population like the Agta. Again, however, these findings need to be interpreted with caution – but clearly this is an avenue to explore in future research.

3.3. Study 3: Turkish parents

3.3.1. Methods

3.3.1.1. Participants

Participants were 26 opportunity-sampled families, resident in Ankara, Turkey (fathers' age range 28-41 years, *M* = 36.46, *SD* = 3.47; mothers' age range 28-41 years, *M* = 34.85, *SD* = 3.45). All participating families had at least one child. The families had previously taken part in one data collection where parents had their facial photographs taken and reported own reproductive outcomes and offspring health outcomes; here, we report reproductive outcomes.

3.3.1.2. Materials and procedure

The study protocol was approved by Durham University Department of Psychology Ethics Committee. All participating parents gave informed consent. Families were not renumerated. Data collection took place in the families' own homes.

Parental facial images. Photographs were taken front-facing. Facial expressions in photographs were generally neutral, but some variation existed within the dataset. Lighting conditions, facial angle and tilt, and image resolution also varied somewhat between parents, and photographs were therefore semi-standardised. Parents were photographed with their hair tied back. One mother reported having had cosmetic plastic surgery on her face and was omitted from analysis.

Reproductive variables. Parents reported their total number of offspring and the number of offspring who had survived until age 5; the proportion of surviving offspring was calculated. However, since only one couple reported having deceased offspring, this variable was extremely highly correlated with total number of offspring ($r = .966$, $p < .001$). It was therefore deemed redundant and was omitted from analysis, leaving just one reproductive variable: total number of offspring. This variable was significantly skewed and was square root-transformed, whereby it approached but did not reach normality.

Rated facial traits. Facial images of parents were cropped and masked along the outline of the faces, excluding the ears and the neck. Images had previously been rated by observers for attractiveness, masculinity (in men) and femininity (in women), and for perceived health, on a scale from 1-7. Image order was randomised. Observers were recruited online; no other information about observer demographics was available. Average ratings were calculated for each face on each trait. Both father's and mother's facial attractiveness were skewed and were therefore log 10-transformed.

3.3.1.3. Statistical analysis

Means and standard deviations for all study variables can be viewed in Table 3.4. Father's facial hair was significantly correlated with rated masculinity ($r = .708$, $p < .001$) but not with the other facial traits. Mum's makeup use was significantly correlated with all three facial traits ($r \geq .665$, $p < .001$ for all). We therefore controlled for these variables by saving the residuals from separate linear regressions with, for fathers, facial hair as the predictor and facial attractiveness as the outcome, and for mothers, makeup use as the predictor and each of the three facial traits as the outcome. The saved residuals were used for analyses. We ran

Pearson’s correlations to test the hypothesis that parental dimorphism and attractiveness predicts greater fertility. As in Study 2, we also checked for quadratic associations.

Table 3.4

Descriptive statistics for all study variables in the Turkish parents (Study 3).

Father’s traits	*N*	Min	Max	Mean	*SD*
Facial masculinity	26	3.64	6.52	5.18	0.81
Facial attractiveness	26	1.54	3.42	2.25	0.45
Facial health	26	3.06	4.89	3.98	0.51
Mother’s traits	***N***	**Min**	**Max**	**Mean**	***SD***
Facial femininity	25	3.21	5.50	4.41	0.68
Facial attractiveness	25	1.68	4.45	2.69	0.82
Facial health	25	2.62	5.78	4.18	0.90
Reproductive variables					
Total number of offspring	26	1	4	1.65	0.75
Proportion of offspring surviving until age 5 (omitted from analysis)	26	0.50	1.00	0.98	0.10

3.3.2. Results

Father’s facial traits were not significantly correlated with each other (all $p > .05$), controlling for age and (for facial masculinity) facial hair, but there was a nonsignificant trend for more facially masculine fathers to be rated as less attractive ($r = -.348$, $p = .081$, $q = .375$). Mother’s facial traits, controlling for age and makeup use, were all strongly positively correlated: femininity and attractiveness $r = .705$, $p < .001$, $q = .011$; femininity and perceived health $r = .741$, $p < .001$, $q = .011$; attractiveness and perceived health $r = .858$, $p < .001$, $q = .011$.

The predictions were not supported: no parental traits significantly predicted a greater number of offspring. Interestingly, all maternal facial traits as well as paternal facial attractiveness were nonsignificantly associated with having *fewer* offspring; the effects for father’s facial attractiveness (r = -.345, p = .084, q = .375) was marginally significant. Thus, for both Agta mothers and for Turkish mothers and fathers, more attractive parents may have *fewer* rather than more offspring, but here we lack the power to draw firm conclusions. We detected no significant quadratic associations.

Table 3.5

Pearson’s bivariate correlations between parental traits and reproduction in the Turkish parents, controlling (where relevant) for parental age, father’s facial hair, and mother’s makeup use (Study 3).

Father’s traits	Total no of offspring
Facial masculinity	.076 *N* = 26
Facial attractiveness	-.345 *N* = 26
Facial health	.076 *N* = 26
Mother’s traits	**Total no of offspring**
Facial femininity	-.332 *N* = 25
Facial attractiveness	-.188 *N* = 25
Facial health	-.244 *N* = 25

Note. No associations were significant even before corrections.

3.4. Study 4: Systematic review and meta-analysis

Next, we report a systematic review of morphological traits predicting reproductive outcomes in women, and a meta-analysis of attractiveness predicting reproductive outcomes in both sexes.

3.4.1. Methods

3.4.1.1. Literature search and study selection

A systematic literature search was conducted in August 2020 using the databases PubMed, PsycINFO, and Web of Science. The following search terms were used: (feminin* OR “sexual dimorphism” OR "sexually dimorphic" OR width-to-height OR fWHR OR “breast size” OR “waist-hip ratio” OR attractive* OR “digit ratio” OR 2d:4d OR “voice pitch” OR “vocal pitch” OR voice OR “fundamental frequency”) AND (“number of offspring” OR “offspring number” OR “number of children” OR “number of grandoffspring” OR “number of grand offspring” OR “offspring health” OR “offspring mortality” OR “mortality of offspring” OR “surviving offspring” OR “offspring survival” OR “reproductive onset” OR “reproductive success” OR “first birth” OR fertil* OR fecund*) AND (human OR woman OR women OR man OR men OR participant*) NOT cancer.

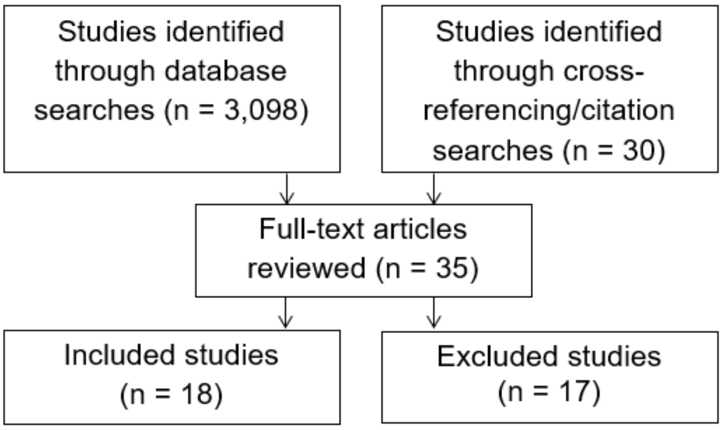
Search alerts were set up to ensure that later publications would be located. Studies were also found through citation searches and cross-referencing, and calls for papers were made on social media. As can be seen in Figure 3.1, a total of 3,098 systematic search results, including duplicates, were retrieved and approximately 30 papers were found through other methods. Abstracts of 78 papers were scanned; 35 of those were read in full. For inclusion, we considered studies which measured at least one of the following predictor traits in *women*: facial femininity, facial attractiveness, pre-pregnancy breast size, pre-pregnancy WHR, 2D:4D, and voice pitch. We did not include studies measuring breast size or WHR after childbirth, since pregnancy and childbirth permanently influence both measures (e.g., Butovskaya et al., 2017). As mentioned above, we also included strength and muscle mass. Most work on offspring viability as a function of maternal morphology has assessed it in relation to maternal height (e.g., Devi et al., 1985; Frisancho et al., 1973; Monden & Smits, 2009; Subramanian et al., 2009; Özaltin et al., 2010). This relationship is strongly influenced by nutritional status (Silventoinen, 2003) rather than reflecting putative maternal quality. The main focus of this thesis is parental quality indicators, and the literature on maternal height predicting offspring viability will therefore not be reviewed here.

In *women*, studies had to measure at least one of the following outcomes: number of offspring/grand-offspring, number of pregnancies, age at first birth, or offspring survival/mortality. In *men*, studies had to include measures of facial attractiveness predicting number of offspring/grand-offspring, age at first birth, or offspring survival/mortality.

Studies had to include adult participants (≥ 17 years of age) and had to analyse men and women separately. If analyses had been conducted both with and without controlling for parental age, we favoured analyses where age had been controlled for. Some effects were converted to Cohen's *d* for ease of interpretation. In total, 18 papers were eligible for inclusion, including 63 observations from 24 samples in 16 countries. Both father's and mother's facial attractiveness had enough observations for us to be able to meta-analyse the effects, whereas mother's morphological body traits did not.

Figure 3.1.

Overview of the systematic literature search.



3.4.2. Results

3.4.2.1. Systematic review of mother's morphological body traits predicting reproductive outcomes

As can be seen in Table 3.6, 9 papers reported associations between maternal traits and reproductive outcomes in a total of 14 different samples. Two papers sampled the same high-fertility Hadza hunter-gatherer population (Apicella et al., 2007; Smith et al., 2017). In that sample, neither mother's upper-body strength/muscularity nor voice pitch significantly predicted number of children born. In contrast, in the high-fertility Himba pastoralists, Atkinson and colleagues (2012) found that women with a higher (i.e., more feminine) voice pitch had significantly more offspring and grand-offspring, whereas women with greater (i.e., more masculine) handgrip strength had more offspring but not significantly more grand-offspring. In all of the remaining papers, primarily using Western samples, reproductive outcomes were measured as a function of 2D:4D. Across 8 out of 24 total observations, women with a more feminine 2D:4D had birthed significantly more children and in three cases, the opposite relationship was found (but for some of the significant relationships, the effect sizes were small enough to be negligible). All of the remaining associations were non-significant.

Table 3.6

Summary of associations between women's morphological body traits and reproductive outcomes.

Author(s) & year	Sample	Trait	Outcome	*N*	Effect
Manning et al., 2000	Community and student sample ≥ 30 yrs, U.K.	R2D:4D	No of offspring	183	+ve: *B* = 5.33
		L2D:4D	No of offspring	183	+ve: *B* = 6.71
	Community sample and university staff ≥ 30 yrs, Germany	R2D:4D	No of offspring	96	+ve: *B* = 6.91
		L2D:4D	No of offspring	96	+ve: *B* = 7.10

	Community sample aged ≥ 30 yrs, Hungary	R2D:4D	No of offspring	39	+ve: *B* = 11.45
		L2D:4D	No of offspring	39	+ve (one-tailed): *B* = 10.04
	Community sample aged ≥ 30 yrs, Poland	R2D:4D	No of offspring	103	Null: *B* = -4.93
		L2D:4D	No of offspring	103	Null: *B* = -2.29
	Community sample aged ≥ 30 yrs, Jamaica	R2D:4D	No of offspring	60	Null: *B* = 6.88
		L2D:4D	No of offspring	60	Null: *B* = 9.32
Manning et al., 2003	Community sample ≥ 23 yrs, U.K.	2D:4D	No of offspring	214	+ve: *r* = .18
	Sugali and Yanadi tribal groups ≥ 18 yrs, India	2D:4D	No of offspring	80	Null: *r* = .21
	Zulus from townships near Durban ≥ 40 yrs, South Africa	2D:4D	No of offspring	132	-ve: *r* = -.25
Vehmas et al., 2006	Teachers and dentists aged 45-63 yrs, Finland	2D:4D	No of pregnancies	489	Null: *r* = .006
Apicella et al., 2007	Hadza, Tanzania	Voice pitch	No of offspring born	48	Null: β = -.111
			Reproductive success	48	Null: β = .058
			No of offspring dead	48	Null: β = .023
			Mortality rate	48	Null: β = .215
Manning & Fink, 2008	Online sample aged ≥ 18 yrs, worldwide	R2D:4D	No of offspring	69173	-ve: *r* = -.026
		L2D:4D	No of offspring	69173	-ve: *r* = -.019
Helle, 2010	Community sample, Finland (born 1946-58)	R2D:4D	Age at first birth (only women who had reproduced)	240	Null: *r* = -.06
		L2D:4D	Age at first birth (only women who had reproduced)	240	Null: *r* = .02
		R2D:4D	No of adult offspring (only women who had reproduced)	240	Null: *r* = .01
		L2D:4D	No of adult offspring (only women who had reproduced)	240	Null: *r* = .04

Atkinson et al., 2012	Himba (Ovahimba), Namibia	Voice pitch	Genetic vector*	54	+ve: $\beta = .292$
			No of living offspring	54	+ve: $\beta = .245$
			No of living grand-offspring	54	+ve: $\beta = .239$
			No of living grand-offspring (controlling for offspring)	54	+ve: $\beta = .168$
		HGS	Genetic vector*	54	+ve: $\beta = .281$
			No of living offspring	54	+ve: $\beta = .327$
			No of living grand-offspring	54	Null: unknown
			No of living grand-offspring (controlling for offspring)	54	Null: unknown
Klimek et al., 2016	Mogielica Human Ecology Study Site, Poland	R2D:4D	No of offspring	298	+ve: $\chi^2 = 13.1$
			Age at first birth	291	Null: $d = 0.15$
		L2D:4D	No of offspring	301	Null: $\chi^2 = 1.63$
			Age at first birth	291	Null: $d = 0.05$
Smith et al., 2017	Same as Apicella et al., 2007	Upper-arm muscle mass/ HGS	No of offspring born	48	Null: $r = .09$
			Reproductive success	48	Null: $r = .18$
			Offspring mortality	45	Null: $r = -.18$

Note. *1*(number of living children) + 1/2 *(number of living grandchildren). HGS = handgrip strength; L2D:4D = left 2D:4D; R2D:4D = right 2D:4D. Since the predicted direction of effects differ between studies depending on measures used, effects showing significantly better reproductive outcomes as a function of increased femininity are indicated in green and as a function of increased masculinity in orange. We do not indicate marginally significant effects as significant here (as is sometimes done in the papers where they were reported). Note that we have not included measures of height here.

3.4.2.2. Meta-analysis of parents’ facial attractiveness predicting reproductive outcomes

3.4.2.2.1. Statistical analysis

As in the previous chapter, we used the *metafor* package (Viechtbauer, 2010) in R 4.1.0 (R Core Development Team, 2021) to perform a meta-analysis of the hypothesis that parental facial attractiveness is associated with better reproductive outcomes. Six studies had measured the relevant association in men and seven in women. Hill and Hurtado (1996) reported that in one of two samples of Ache women, facially attractive women had significantly more offspring (β = .07) but not more surviving offspring; however, sample sizes were not reported for these observations, why they had to be omitted from the analysis. Silva and colleagues (2012) also reported reproductive outcomes as a function of attractiveness, but effect size measures in the article were ambiguous and the article was therefore excluded.

To the published effects where effect- and sample sizes were available, we also added effects sizes from Studies 2 and 3 above. In total, this gave us nine effects from seven different samples in men, and 14 effects from seven samples in women. All studies included in the analysis can be viewed in Table 3.7. We used Pearson’s *r* as our effect size measure; effect sizes are presented as Pearson’s *r*. Outcome measures that were expected to show negative relationships with parental attractiveness (such as age at first birth and offspring mortality) were reverse coded, so that all predicted relationships between parental facial traits and reproductive outcomes were positive. Effects were clustered by sample and analyses were conducted using random-effects models.

Table 3.7

All studies included in the meta-analysis (Study 4).

Authors	Year	Sample	Sample location	Low or high fert.	*N*
Boothroyd et al.	2017	Agta	Philippines	High	65 (M)
		Maya	Belize	High	23-35 (M)
Jokela	2009	Wisconsin Longitudinal Study (born 1937-40)	U.S.	Neither	997 (M), 1244 (F)
Kalick et al.	1998	Intergenerational Studies (born 1920-29)	U.S.	High	116 (M), 127 (F)
Lidborg et al.	2021	Community sample	Turkey	Low	26 (M), 25 (F)
		Agta	Philippines	High	94-105 (F)
Pawlowski et al.	2008	Rural sample	Poland	High	47 (F)
Pflüger et al.	2012	Rural community sample	Austria	High	40 (F)
		Rural community sample	Austria	Low	46 (F)
Prokop & Fedor	2011	Friends and family of students	Slovakia	Low	499 (M)
Prokop et al.	2010	Friends and fathers of students	Slovakia	Low	245 (M)

Note. F = females; M = males. The sample in Jokela (2009) reproduced during the 1960s when the U.S. transitioned from high to low fertility and the sample was therefore coded as neither fertility type.

3.4.2.2.2. Results

Our hypothesis was not supported in either sex: there were no significant associations between father's or mother's facial attractiveness and reproductive outcomes (Table 3.8). In addition to analysing low and high fertility samples together, we also analysed these subsamples separately: no effects were significant. Effect sizes for fathers were around zero for both sample types. For mothers, the effect size was zero in low fertility samples, but stronger in high fertility samples. See Appendix I for funnel plots.

Table 3.8

Parental facial attractiveness predicting reproductive outcomes. Pearson's r (95% CI); p value for meta-analytic effect, q value (correcting for multiple comparisons); number of observations (k), and 'unique' participants (n); test for heterogeneity (Q), p value for heterogeneity. Statistically significant meta-analytic associations are bolded if still significant after controlling for multiple comparisons.

	Father's facial attractiveness	Mother's facial attractiveness
Reproductive outcomes: All samples	*r* = .046 (-0.025, 0.117), *p* = .202, *q* = .575 *k* = 9, *n* = 1,983 Q(df = 8) = 11.904, *p* = .156	*r* = .080 (-0.043, 0.203), *p* = .203, *q* = .575 *k* = 14, *n* = 1,634 Q(df = 13) = 24.626, *p* = .026
Reproductive outcomes: Low fertility samples	*r* = -.022 (-0.259, 0.215), *p* = .854, *q* = .895 *k* = 3, *n* = 770 Q(df = 2) = 10.271, *p* = .006	*r* = .003 (-0.127, 0.133), *p* = .962, *q* = .895 *k* = 4, *n* = 1,315 Q(df = 3) = 2.864, *p* = .413
Reproductive outcomes: High fertility samples	*r* = .030 (-0.085, 0.145), *p* = .610, *q* = .818 *k* = 5, *n* = 216 Q(df = 4) = 1.198, *p* = .878	*r* = .156 (-0.031, 0.347), *p* = .106, *q* = .413 *k* = 10, *n* = 319 Q(df = 9) = 21.273, *p* = .012

Note. *k* = number of observations; *n* = number of unique participants; Q = Cochran's Q test of heterogeneity; *q* = *q* value.

3.5. Discussion

In both men and women, sexually dimorphic traits are commonly argued to signal good mate quality, frequently indexed as higher immunity and/or higher fertility. In this chapter, we reviewed evidence for these claims in women. Overall, to date, there is not sufficient or conclusive evidence that women's morphological traits signal either quality measure. Furthermore, to test the prediction that attractiveness signals better biological fitness – which is key to demonstrate that attractiveness preferences are adaptive – we also analysed the relationships between facial attractiveness and reproductive outcomes in a sample of Agta hunter-gatherer women (Study 2) and one sample of Turkish parents (Study 3). Whereas facially attractive Agta women had significantly more offspring surviving until age 5, facially healthy-looking women reported *fewer*, rather than more, births. Neither association remained significant after we corrected for multiple comparisons. In the Turkish parents, we detected no significant relationships – but parents' facial traits were generally associated with having *fewer* offspring, in contrast to predictions. Lastly, we meta-analysed the associations between facial attractiveness and reproductive outcomes in both men (k = 9) and women (k = 14). The associations were not significant in either sex, but it is noteworthy that effect sizes for women were larger when restricting the analysis to high fertility populations. Thus, while our analysis is limited in that samples were few, it does suggest that men's facial attractiveness is not associated with better fitness outcomes and is not under present selection, but future research should investigate whether it may be under selection in non-contracepting women. If this association holds true in larger samples, future research should attempt to answer whether this is mediated simply by more attractive being more likely to marry – overall and/or at a younger age – or whether they are indeed more fecund compared to less attractive women. In the Agta women, we did not observe that attractive women started reproducing earlier, but Kalick and colleagues (1998) reported that more attractive U.S. women did marry earlier. In a similar vein, Rhodes and colleagues (2005) found that more facially feminine Australian women reported earlier first sexual intercourse and a greater number of long-term relationships (but see also Pawlowski et al., 2008, and Foo et al., 2017c, for null findings). If attractive women *do* marry – and thereby start reproducing – earlier, this is problematic for the claim that human attractiveness preferences evolved to secure the acquisition of a high quality (i.e., in women, fertile) mate as it inverts cause and effect. This would also indicate that facial attractiveness is under selection in women.

In the review, the majority of effects for 2D:4D were positive, suggesting that digit ratios – which are often claimed to reflect prenatal hormone exposure (Manning, 2002) – might cue reproductive capacity in women. Given that this finding is based on few samples, however, future research will need to show whether or not this association is robust.

Our meta-analysis showed no evidence suggesting that men’s facial attractiveness increases biological fitness. While this is based on few samples, it can be noted that effect sizes were close to zero across populations. Overall, this does not suggest that facial attractiveness in men is under selection. This challenges the view that humans are attracted to features associated with better mate quality. While it is important to note that facial attractiveness and facial masculinity are not consistently correlated in men, in the meta-analysis in Chapter 2, men’s facial masculinity was not significantly associated with either mating or reproductive outcomes, whereas body masculinity was. Such findings *may* imply that facial features, while doubtless highly salient social stimuli, in general do not communicate mate quality in the way that is commonly assumed (perhaps not in either sex).

It is important to note that methods were rather heterogenous across studies. For instance, in some studies, participants’ attractiveness was assessed from participant-provided (i.e., non-standardised) photographs from their youth (pre-reproduction), whereas in other studies the participants were photographed by experimenters, sometimes late in life, and participant age at the time of being photographed therefore varied considerably between studies. It cannot be ruled out that this may have influenced the pattern of results, and such methodological differences should ideally be addressed in the future. We were also required to exclude a number of studies due to women’s body traits being measured post-reproduction. Ideally, future research will use longitudinal methods to assess the association between facial and body traits – measured pre-marriage/pre-reproduction – and lifetime reproductive output. This would also enable us to determine whether a relationship between women’s attractiveness and fertility is mediated by age at/likelihood of marriage.

Chapter 4.

Parental morphology and attractiveness as indicators of offspring quality

---

4.1. Introduction

4.1.1. Predictions based on the ICHH

A number of key predictions follow from the ICHH, including – but not limited to – the following: 1. Testosterone should be immunosuppressive; 2. Men high in testosterone-dependent, i.e. masculine, traits should show better immunity/health; 3. The offspring of masculine men should inherit their father's immunocompetence; 4. Highly masculine men should, due to their higher quality, be able to acquire better-quality mates (i.e. healthier and/or more fertile, as noted in the previous chapter typically indexed in women by increased attractiveness/femininity), which might mediate the quality of these men's offspring. In this chapter, we first review evidence for the two first predictions, before presenting the results of two empirical studies where we tested the third and fourth of these predictions: that paternal traits should predict better offspring health and viability, controlling for maternal traits.

In Chapter 2 (Study 1), we showed that men higher in body masculinity show greater mating success than less masculine men, where one aspect of mating success was age at sexual onset. If traits that facilitate men's mating success are heritable, it can be expected that more masculine men should also father offspring with greater mating success. Here, we therefore also tested whether paternal masculinity predicted earlier sexual activity in offspring.

4.1.2. The association between men's testosterone levels and health/immunity

In terms of the first prediction, evidence for an association between testosterone and immunity is tentative. Meta-analyses using data from non-human animals provide, at best, weak support. When combining observations across species, Roberts and colleagues (2004) found that testosterone significantly but weakly suppressed immune functioning. When the authors controlled for non-independence of studies and analysed species separately, however, the relationship was no longer significant in mammals or birds, remaining significant only in reptiles. More recently, Foo and colleagues (2017a) reported that testosterone showed suppression effects on immunity in experimental but not correlational studies. If testosterone suppresses the immune system, it follows that females should show stronger immune responses than males: an analysis by Kelly and colleagues (2018) found that this was indeed the case. As in the analysis by Roberts and colleagues, however, the overall effect was weak, and was no longer significant after controlling for phylogeny.

Evidence from humans also illustrates the complexity of the relationship between testosterone and immune functioning. In a meta-analysis, Guerra-Silveira and Abad-Franch (2013) observed an overall male bias in infectious disease prevalence in humans, as would be expected based on the ICHH. It should be noted here, however, that increased body size – which, as previously mentioned, is seen in human males – is typically associated with an increased parasite burden independently of immunosuppression caused or influenced by sex hormones (McDade, 2003). A review by Boothroyd and colleagues (2013) showed a mixture of contradictory – where testosterone was associated with increased as well as decreased disease prevalence – and null findings. Since the publication of their review, both positive, negative, and nonsignificant associations have been reported between testosterone and a number of immunity measures. High testosterone has, for example, variously been linked to stronger (Pawlowski et al., 2018) and weaker post-vaccination antibody responses (Furman et al., 2014) as well as showing no relationship (Skrinda et al., 2014). Pawlowski and colleagues (2018) also observed that men with high testosterone levels showed stronger immune responses as indicated by seroconversion and the antibody immunoglobulin G (IgG) (but it should be noted that in their study, 9 out of 12 measures showed no significant relationship with testosterone). Other researchers have also found testosterone to be positively associated with secretory immunoglobulin A (SIgA: another type of antibody) (Arnocky et al., 2018; Gettler at al., 2014; Hodges-Simeon et al., 2019; Hodges-Simeon et al., 2020). Trumble and colleagues (2016) reported evidence of immunosuppressive effects of testosterone on some but not all measured cytokine responses in men. Furthermore, in the coronavirus disease 2019 (COVID-19) pandemic, men show overall higher incidence of severe illness and higher mortality rates compared to women (Pradhan & Olsson, 2020). Interestingly, the role of testosterone seems to constitute a 'double-edged sword' in COVID-19, whereby both high and low levels of testosterone have been linked to an increased risk of requiring intensive care and higher mortality (e.g., Salonia et al., 2021; Traish & Morgentaler, 2021; Younis et al., 2021). Such correlational relationships should not be interpreted as suggesting causal mechanisms, however.

Overall, then, the evidence for a global immunosuppressive role of testosterone is limited, and the evidence to date rather points to testosterone being immunomodulatory. Under this notion, testosterone may upregulate some aspects of immune functioning while downregulating others (e.g., Hodges-Simeon et al., 2019; Muehlenbein & Bribiescas, 2005; Nowak et al., 2018; Trumble et al., 2016).

4.1.3. The association between men's morphological traits and health/immunity

Evidence for the second prediction, that masculinity should be associated with better health, is similarly mixed (Table 4.1). Early studies indicated that there may be an association between men's facial masculinity and improved health outcomes. In these studies, facial masculinity was typically measured either through observer ratings or through geometric morphometric analysis) a. Men with increased rated facial masculinity have been found to show improved health, indexed by childhood and adolescent medical examinations and health histories (Rhodes et al., 2003). Using measured facial masculinity, Thornhill and Gangestad (2006) reported that more masculine men self-reported fewer respiratory infections and less frequent antibiotic use (Thornhill & Gangestad, 2006). In contrast, Boothroyd and colleagues (2013) observed no significant associations between rated facial masculinity and health, and while morphometrically masculine men in their study reported fewer respiratory infections in the past, they reported *more* infections at follow-up. In neither of the two aforementioned studies did facially masculine men differ from less masculine men in their incidence of gastroenteritis (gastrointestinal infection).

In terms of more direct measures of immune functioning, men with increased rated facial masculinity have been observed to display stronger antibody responses (Rantala et al., 2013; Skrinda et al., 2014) but no significant difference in cytokine response or C-reactive protein (Phalane et al., 2017). Null findings have also been reported between rated facial masculinity and composite immune measures indexed by salivary antibacterial capacity, overall bacterial immunity, bacterial suppression capacity, and lysozyme activity (Foo et al., 2017b) and between both rated and morphometric masculinity and MHC heterozygosity (Lie et al., 2008; Zaidi et al., 2019). Foo and colleagues (2020) rightly argued that facial dimorphism – which develops primarily in adolescence – should reflect adolescent rather than adult health. They consequently measured a wide range of immune responses at a single timepoint in adolescence as a function of adult morphometric facial dimorphism. Men’s facial masculinity significantly predicted a limited set of the measured immunity measures; those that were significant largely reflected elevated allergic responses and increased antibacterial and cellular immunity.

The prediction that facial masculinity should signal good genetic quality is intertwined with the notion that women should be attracted to men's masculine traits; thus, masculinity and attractiveness should be correlated, which is not consistently the case (e.g., Boothroyd et al., 2017; Foo et al., 2017b; Mitchem et al., 2013; Rantala et al., 2013; Skrinda et al., 2014). Similarly to facial masculinity, men's facial attractiveness has variously been linked to better health outcomes (Boothroyd et al., 2013; Henderson & Anglin, 2003; Lie et al., 2008; Rantala et al., 2012; Rantala et al., 2013a; Roberts et al., 2005; Shackelford & Larsen, 1999) and showing no relationship (Boothroyd et al., 2013; Foo et al., 2017b; Hume & Montgomerie 2001; Kalick et al., 1998; Phalane et al., 2017; Thornhill & Gangestad, 2006; see also Zebrowitz & Rhodes, 2004, for a re-analysis of the data reported by Kalick et al., 1998) – different health measures oftentimes showing conflicting relationships with attractiveness within the same study.

Compared to facial traits, other masculine traits have received relatively less attention in the literature. Men whose voices are rated as more dominant-sounding and deeper appear to show better health (Albert et al., 2021a; Arnocky et al., 2018), but no significant associations have been reported between vocal attractiveness and the strength of antibody responses (Skrinda et al., 2014) or self-reported health (Albert et al., 2021a). In another study, men's self-reported muscularity and strength, but not their height or 2D:4D, were significantly related to better self-reported health (Ellis & Hoskin, 2020). To our knowledge, no other research has assessed the relationship between health and non-facial masculine traits in men. Height is the only exception, in that it has been studied extensively in relation to health-related outcomes and such relationships have been reviewed elsewhere (e.g., Stefan et al., 2016). In terms of immune responses, men's increased height may be associated with better immunity, indexed by a stronger antibody response (Krams et al., 2014; Skrinda et al., 2014) and MHC heterozygosity (Zaidi et al., 2019), at least up to a height of 185-188 cm (Krams et al., 2014; Skrinda et al., 2014). However, Pawlowski and colleagues (2014) found no significant relationship between height and pathogen colonisation in men.

Overall, then, it is clear that the putative positive association between men's morphological traits and improved health/immunity outcomes is neither strongly nor consistently supported by empirical evidence. The majority of research done to date has focussed on facial traits, and not enough attention has been focussed on testing health as a function of masculinity in men's bodies. Given that, in humans, traits such as muscle mass and physical strength show higher degrees of sexual dimorphism than facial traits do and, unlike facial traits, continue to be highly metabolically costly past trait development in puberty, there is likely a stronger theoretical basis for hypothesising a relationship between such traits and health/immunity – be it due to such traits being reflective of good genes/immunocompetence or due to such traits being condition-dependent.

Table 4.1

Summary of associations between male dimorphism/attractiveness and health/immunity outcomes.

Facial traits

Author(s) & year	Sample	Trait	Outcome	Notes	Effect
Kalick et al., 1998	Inter-generational Studies, U.S. (born 1920-29)	Attr.	Childhood/ adolescent health	Attr. rated at age 17-18 yrs. Health rated by physicians through clinical exams and histories at 11-18 years	Null
			Middle adulthood health	Health based on a medical exam and history at 30-36 years	Null
			Later adulthood health	Heath rated by physicians at 58-66 years	Null
Shackelford & Larsen, 1999	University students, U.S.	Attr.	Headache	Self-reported	Null
			Runny or stuffy nose	Self-reported	-ve: r = .25
			Nausea or upset stomach	Self-reported	Null
			Muscle soreness or cramps	Self-reported	Null
			Sore throat or cough	Self-reported	Null
			Backache	Self-reported	Null

			Jitteriness or trembling	Self-reported	Null
			Cardiovascular health	Measured by cardiac recovery time	Null
Hume & Montgomerie, 2001	University students, Canada	Attr.	Common health problems	Health assessed by composite of fractures, various diseases, and cosmetic or reconstructive surgeries	Null
Rhodes et al., 2003	Same as Kalick et al., 1998	Rated masc.	Childhood/ adolescent health	Health rated by physicians through clinical exams and histories at 11-18 years	+ve: $r = .17$
Zebrowitz & Rhodes, 2004 (re-analysis of Kalick et al., 1998)	Same as Kalick et al., 1998	Attr. above median	Childhood/ adolescent health	Attr. assessed in adolescence (men and women analysed together throughout)	Null
		Attr. below median	Childhood/ adolescent health	Attr. assessed in adolescence	Null
		Attr. above median	Childhood/ adolescent health	Attr. assessed in adulthood	Null
		Attr. below median	Childhood/ adolescent health	Attr. assessed in adulthood	+ve: $r = .30$
Henderson & Anglin, 2003	High school students, Canada (born ~1907-10)	Attr.	Longevity	Attr. assessed through photographs from high school yearbooks, ~17 years	+ve: $r = .34$
Roberts et al., 2005	University students/staff, U.K.	Attr.	MHC heterozygosity	Independent samples *t* test between hetero- and homozygous men	+ve: $t(90) = 2.29$
Thornhill & Gangestad, 2006	University students, U.S.	Measured masc.	Number of respiratory infections	Illnesses measured in the past 3 years, self-reported	-ve: $r = -.19$
			Days of respiratory infections		-ve: $r = -.16$
			Number of gastrointestinal infections		Null
			Days of gastrointestinal infections		Null
			Antibiotic use		-ve: $r = -.21$
		Attr.	Number of respiratory infections	Illnesses measured in the past 3 years, self-reported	Null
			Days of respiratory infections		Null
			Number of gastrointestinal infections		Null
			Days of gastrointestinal infections		Null
			Antibiotic use		Null
Lie et al., 2008	University students, Australia	Rated masc.	Mean MHC heterozygosity		Null
			Standardised mean genetic distance between alleles		Null
		Attr.	Mean MHC heterozygosity		+ve: $r = .33$

			Standardised mean genetic distance between alleles		Null
Rantala et al., 2012	Students, Latvia	Attr.	Hepatitis B antibody response		+ve: *r* = .50
Boothroyd et al., 2013	University students, U.K.	Morph. masc.	Bouts of colds/flu	Illnesses measured in the past 3 years, self-reported	-ve: *r* = -.338
			Bouts of gastroenteritis		Null
			Antibiotic use		Null
			Days with colds	Illnesses measured at follow-up, self-reported	Null
			Days with flu		+ve: *r* = .318
			Days with gastroenteritis		Null
			Antibiotic use		Null
		Rated masc.	Bouts of colds/flu	Illnesses measured in the past 3 years, self-reported	Null
			Bouts of gastroenteritis		Null
			Antibiotic use		Null
			Days with colds	Illnesses measured at follow-up, self-reported	Null
			Days with flu		Null
			Days with gastroenteritis		Null
			Antibiotic use		Null
		Attr.	Bouts of colds/flu	Illnesses measured in the past 3 years, self-reported	-ve: *r* = -.330
			Bouts of gastroenteritis		Null
			Antibiotic use		Null
			Days with colds	Illnesses measured at follow-up, self-reported	Null
			Days with flu		Null
			Days with gastroenteritis		Null
			Antibiotic use		-ve: *r* = -.350
Rantala et al., 2013	Same as Rantala et al., 2012	Rated masc.	Hepatitis B antibody response		+ve: *r* = .47
		Attr.	Hepatitis B antibody response		+ve: *r* = .43
Skrinda et al., 2014	Same as Rantala et al., 2012	Rated masc.	Hepatitis B antibody response		+ve: *r* = .62
		Attr.	Hepatitis B antibody response		-ve: *r* = -.21
Phalane et al., 2017	University students, South Africa	Rated masc.	Cytokine component		Null
			C-reactive protein		Null
		Attr.	Cytokine component		Null
			C-reactive protein		Null
Foo et al., 2017b		Rated masc.	Bacterial killing capacity/ overall bacterial immunity		Null

	University students, Australia		Bacterial suppression capacity/ lysozyme activity		Null
		Attr.	Bacterial killing capacity/ overall bacterial immunity		Null
			Bacterial suppression capacity/ lysozyme activity		Null
Zaidi et al., 2019	Sample of European ancestry	Morph. masc.	MHC heterozygosity		Null
Foo et al., 2020	The Western Australian Pregnancy Cohort (Raine) Study	Morph. masc.	Immunity composite 9 (allergies)	Masculinity measured at 22 years, immunity at 14 years. Immunity measures across 16 composites included haematology, antibodies, cytokine responses, inflammatory markers	+ve: $r = .15$
			Immunity composite 15 (allergies, antibacterial immunity, cellular immunity)		+ve: $r = .15$
			14 separate immunity composites		Null
			Bodily traits		
Rantala et al., 2013a	Same as Rantala et al., 2012	Rated body masc.	Hepatitis B antibody response		Null
		Body attr.	Hepatitis B antibody response		+ve: $r = .47$
Skrinda et al., 2014	Same as Rantala et al., 2012	Rated vocal masc.	Hepatitis B antibody response		Null
		Rated vocal attr.	Hepatitis B antibody response		Null
Arnocky et al., 2018	University students, Canada	f_0	sIgA		-ve: $r = -.21$
			General health status	Self-reported	Null
			Current health status	Self-reported	Null
			Future health	Self-reported	Null
			Infectability	Self-reported	+ve: $r = .17$
Ellis & Hoskin, 2020	University students, Malaysia	Strength	Self-reported health		+ve: $\rho = .621$
		Musc-ularity	Self-reported health		+ve: $\rho = .281$

		2D:4D (inverted)	Self-reported health		Null
	University students, U.S.	Strength	Self-reported health		+ve: ρ = .460
		Muscularity	Self-reported health		+ve: ρ = .408
		2D:4D (inverted)	Self-reported health		Null
Albert et al., 2021a	University students, Canada	Vocal attr.: LT	Self-reported health	Vocal attractiveness rated by women	Null
		Vocal attr.: ST	Self-reported health	Vocal attractiveness rated by women	Null
		Vocal physical dom.	Self-reported health	Vocal dominance rated by men	+ve: *b* = 0.86, *t* = 3.49
		Vocal social dom.	Self-reported health	Vocal dominance rated by men	+ve: *b* = 0.62, *t* = 3.45

Note. Dom = dominance; f0 = fundamental frequency (voice pitch); LT: long-term relationship context; MHC = major histocompatibility complex; sIgA = salivary immunoglobulin-A; ST = short-term relationship context. Since the predicted direction of effects differ between studies depending on measures used, effects showing significantly better health as a function of *increased masculinity* are indicated in green and worse health in orange. We do not indicate marginally significant effects as significant here (as is sometimes done in the papers where they were reported). Note that we have not included measures of height here.

4.1.4. The association between men's testosterone levels/morphological traits and offspring health/viability

If masculinity-cued immunity is heritable, masculine men should produce healthier, more viable offspring. In spite of being a key prediction of the ICHH, this hypothesis has barely been tested. In our meta-analysis in Chapter 2, we showed that men's body masculinity (primarily measures of strength and muscularity) was significantly predictive of improved reproductive outcomes, but offspring mortality (largely measured as a function of men's height) showed no effect. However, there are indications that a putative relationship between paternal traits and offspring quality may not be linear. Boothroyd and colleagues (2017) reported a curvilinear relationship between paternal facial masculinity and offspring mortality in the Agta of the Philippines and the Maya of Belize, whereby fathers whose faces were rated as moderately masculine had the lowest offspring mortality. Similarly, in Bondongo fisher-farmers in the Republic of Congo, fathers high in testosterone were considered by peers to be better providers – but it was men with intermediate testosterone levels who were ranked as the best caregivers and as having the healthiest children (Boyette at al., 2019). These men also had children in the best energetic condition. These findings should, however, be viewed in light of the fact that *i.* subjective health and energetic condition are not necessarily measures of *immunity* per se, and *ii.* testosterone is reactive, declining in men when they enter fatherhood; the greatest decline being observed in men engaging more in childcare (reviewed in Gray et al., 2020). This makes it challenging to deduce the direction of the relationship between paternal testosterone levels and offspring condition. To date, no research has tested the association between paternal traits and offspring quality, controlling for maternal traits.

4.1.5. The present chapter

To summarise the literature reviewed in this and the previous chapter, hypotheses claim that sexually dimorphic and attractive traits in both men and women signal mate quality, commonly indexed by superior health/immunity and/or fertility; the former being a particularly common claim in men and the latter in women. Despite the prevalence of such claims, the empirical support is oftentimes weak, inconsistent, or lacking completely. Importantly, most research to date has focussed on facial traits in both men and women. Assessing quality in relation to body traits is an understudied area in both sexes.

A key prediction of hypotheses positing that dimorphism/attractiveness indexes heritable quality (regardless of exactly what that quality measure is claimed to be) is that dimorphic/attractive parents should pass on their superior quality to their offspring. This prediction is, to date, severely undertested. Here, we therefore tested the hypothesis that parental traits positively predict better offspring quality in two studies. In Study 5, we tested the association between parental traits (facial traits: facial masculinity/femininity, attractiveness, perceived health, and physical dominance, in addition to the anthropometric measurements handgrip strength and height,) and offspring health in a British sample (pre-registered analysis: https://osf.io/myz2j). Offspring were tested at three timepoints: once in late childhood and twice as teenagers. In Study 6, we report the relationship between parental facial traits (facial masculinity/femininity, attractiveness, and perceived health) and offspring health in the Turkish sample whose reproductive outcomes were reported in Study 3 from Chapter 3. Following from our meta-analytic findings that masculine men show increased mating success, in Study 5 we also tested the prediction that offspring of more dimorphic/attractive parents should show earlier sexual onset, indexed by earlier interest in romantic relationships and sexual activity.

4.2. Study 5: British families

4.2.1. Methods

4.2.1.1. Participants

Tees Valley Baby Study sample. The first subset of participating families was from the longitudinal Tees Valley Baby Study (TVBS). The TVBS families (initial *N* = 206 offspring) originally consisted of mother-offspring dyads, before fathers were invited to take part in later data collections. The families were first recruited into the study in 2002-03, when the focal offspring were infants, through mother and baby groups and through healthcare professionals. The families have participated in intermittent data collections since then (reported in e.g., Meins et al., 2011; Vukovic et al., 2015). The families originate from the Tees Valley area in northeast England and are socioeconomically diverse. The majority of the families are white.

In this chapter, we report data collected during three rounds of larger lab-based data collections – in 2011, 2012, and 2018 – and one telephone-based follow-up data collection in 2019. All families, for whom contact details existed, were invited to participate in the three lab-based data collections; only families who had participated in 2018 were contacted for the 2019 telephone-based follow-up. The exact sample of participating families thus varied between data collections. Also note that, although both parents were invited to participate with their offspring, in many cases only one parent chose to participate.

Parents were photographed in 2011 and in 2018; in 2018, parents also had their anthropometric measurements taken. In 2012 and 2018, health data for the offspring were collected from parents. In the telephone-based data collection in 2019, offspring health data was collected from either the parents or from the offspring themselves. Offspring's romantic interest was measured in 2012 and 2018; sexual activity in 2018.

In 2011, 107 parents were photographed (45 fathers). In 2012, when the offspring were 10 years old, 88 offspring (44 male) took part. In 2018, when the offspring were 16 years of age, 81 offspring (38 male, including one pair of twins whose data were averaged in all analyses, leaving a total sample of 80 offspring, 37 of them male), and 104 parents (29 fathers) participated. In the 2019 follow-up, 73 families responded (36 male offspring). Note that some parents were photographed in both 2011 and 2018, in which case the best quality image was selected. Information about parents' age was collected for parents participating in 2018 but not for all the parents in 2011; therefore, for facial traits, parental age in photograph could not be controlled for. However, we did control for whether parents were photographed when their offspring was 9 or 16 years old. Parental age was available for all parents that had their anthropometric measurements taken, since this was done in the 2018 data collection.

Table 4.2

Types of data collected and total sample sizes for each round of data collection (Study 5).

	Photographs (parents)	Anthropometric measurements (parents)	Health (offspring)	Sexual activity (offspring)
2011	√	×	×	×
Male	45			
Female	62			
Total	107			
2012 (children)	×	×	√	√
Male			44	50
Female			44	49
Total			88	96
2018 (teenagers)	√	√	√	√
Male	28	28	37	37
Female	65	75	42	43
Total	93	103	79	80
2019 (follow-up)	×	×	√	×
Male			36	
Female			37	
Total			73	
University (teenagers)	√	√	√	√
Male	13	13	4	8
Female	11	11	14	17
Total	24	24	18	25

University sample. The second subset of families was recruited from university applicants attending post-offer ‘Open Days’ in the Psychology department at Durham University. Health and/or sexual onset data was collected for 25 offspring (8 male, age range 17-19 years, *M* = 17.80, *SD* = 0.58). Thirteen fathers (age range 40-66 years, *M* = 52.46, *SD* = 6.77), and 11 mothers (age range 44-56 years, *M* = 49.27, *SD* = 3.20) were photographed and had their anthropometric measurements taken. Information about socioeconomic status was not collected, but it can be noted that Durham University has a comparatively high intake of privately educated students (Higher Education Statistics Agency, 2021). The sample was predominantly white.

After two non-biological fathers were excluded and the twins' data had been averaged, across the two subsamples a total of 68 fathers and 107 mothers were photographed, and anthropometric measurements taken of 41 fathers (age range 40-66, *M* = 49.98, *SD* = 5.69) and 86 mothers (age range at time of measurement 34-57, *M* = 46.65, SD = 4.72). Offspring outcomes were available for a total sample of 95 children (48 male) and 105 teenagers (45 male). Table 4.2 gives an overview of data collected at the different timepoints and corresponding final sample sizes. Note that exact sample sizes differed between analyses.

4.2.1.2. Materials and procedure

All study protocols were granted ethical approval by the Durham University Department of Psychology Ethics Committee.

In each round of the lab-based data collections, each family visited the lab for one session which was part of a larger data collection where several tasks were not relevant to this study. Families were renumerated between £5 and £30 per visit, depending on the length of the study protocol. Offspring and parents gave informed consent separately. For offspring aged under 16, consent was given in writing by parents and verbally by children, whereas teenagers gave written consent. For the 2019 telephone-based follow-up of the TVBS families, parents had given written informed consent to be contacted for the follow-up when they visited the lab in 2018. For all data collections, both offspring and parents were free to omit any parts of the study protocol they wished.

Parental facial images. Standardised, front-facing photographs were taken of parents using a 12-megapixel Canon DSLR camera with bilateral flash. Parents were photographed in a sitting position against a neutral background and with a neutral facial expression. Their hair was tied back and, if wearing makeup, they were asked to remove it if comfortable doing so. Makeup use was subsequently controlled for in analyses; so was fathers' facial hair.

Parental anthropometric measurements. A wall-mounted tape measure was used to record parents' barefoot height to the nearest millimetre. Height was measured once. Handgrip strength was measured using a hydraulic Takei Hand Grip Dynamometer Analogue 5001. Parents were asked to squeeze the dynamometer as hard as they could for five seconds; three attempts were made with each hand and the strongest attempt was recorded. The same experimenter took the measurements for the majority of the families, except for a total of 10 families which were measured by another experimenter who had been instructed by the first. Measurements for both height and handgrip strength were normally distributed.

Offspring health variables. One parent (usually the mother) from each family reported offspring health, indexed by autoimmune illnesses and incidence of common illnesses (see Table 4.4 for all health variables) experienced in the last three months and in the last three (in children) or four (in teenagers) years, respectively (Table 4.4 and Appendix J). These measures have previously been used by Thornhill and Gangestad (2006) and by Boothroyd and colleagues (2013). Health questionnaires were completed in private. In two instances, both parents of a family filled out the health questionnaire; for consistency, we retained the answers given by the mother. In the 2019 telephone-based follow-up of the TVBS families, an experimenter asked either the offspring him-/herself or a parent the health questions over the phone, after ensuring that it was an appropriate time to do so. Here, health questions asked how many bouts of illness the focal offspring had experienced since the family had visited the lab in 2018. The amount of time in between the lab visit and follow-up varied across families between a little over four to nearly six months, and this was subsequently controlled for in the analyses.

Autoimmune illnesses were combined as a composite variable. In the pre-registration, we stated that the other health variables would be analysed both separately in addition to as a common illness composite (separately for illnesses measured in the past three months and in the past three/four years). Nearly all health variables were skewed and, where possible, were log 10- or square root-transformed to normality prior to analysis; too severely skewed variables (particularly illness measured in the past three months) had to be omitted from analyses. Some of the health variables were therefore only analysed as composite measures. Gastrointestinal symptoms due to lactose intolerance were not included in the gastroenteritis variable; likewise, antibiotics prescribed for non-illness related conditions, such as acne, was not included in antibiotic use. Health questions were open-ended, and participants sometimes gave answers that did not constitute one integer. In those cases, we coded answers in the following ways:

- Answer comprised two consecutive numbers, such as ‘3 or 4’: coded as the midpoint between those two values, i.e. 3.5.
- ‘Numerous’: coded as the maximum score for the sample on that measure.
- ‘Quite a few’: coded as the midpoint between the average and maximum score for the sample on that measure.
- ‘A couple’: coded as 2.
- Answers in excess of a certain value, e.g. ‘12+’: coded as 12.

Offspring sexual onset. The two sexual onset variables were completed in private by the offspring themselves. Romantic interest was measured both in children and in teenagers and indexed by two items: 'I have started to really like the idea of having a boyfriend/girlfriend (e.g., ‘going out’, holding hands, kissing)' and ‘I have 'fancied' or had a crush on someone’. Offspring responded on 4-point scales, ranging from 0 (‘not at all’) to 3 (‘very much’). Responses to the two questions were summed to form a romantic interest composite measure.

Offspring sexual activity was only measured in teenagers, and was indexed by nine items adapted from The Adolescent Sexual Activity Index (Hansen et al., 1999; Appendix K). Due to the age of the respondents, the items only asked about experiences that could be considered precursors to but not *actual* sexual activity (‘Have you engaged in any of the following activities with a romantic partner [e.g., a boy/girlfriend or someone you've ‘pulled’]?’, e.g., ‘kissing’ or ‘cuddling’). The questions were graded from 1 (‘hugging’) to 9 (‘being naked together’), and the ‘highest’ question that had received a ‘yes’ response was used for each respondent. Both romantic interest and sexual activity were slightly skewed. Romantic interest was square root-transformed in the child dataset.

Control variables. All control variable questionnaires were completed in private. In 2012, TVBS parents filled out the Hollingshead Four-Factor Index of Socioeconomic Status (SES; Hollingshead, 1979; Appendix L), adapted for use in the U.K. This index measures parental SES across four domains: marital status, employment status, education, and occupation; responses were summed and higher values thus denote higher SES. SES was slightly skewed and was log 10-transformed to approach normality.

In all the lab-based data collections, families also completed the Pubertal Development Scale (PDS; Petersen et al., 1988; Appendix M) and the Holmes and Rahe Stress Scale for Non-Adults (Holmes & Rahe, 1967; Appendix N). For the TVBS families, parents filled out these scales in 2012; in 2018, teenagers completed the scales themselves. Likewise, teenagers in the university sample completed both scales themselves.

The PDS included five questions in total. Three questions applied to both boys and girls and asked whether they had experienced e.g., body hair growth and skin changes, and two questions were specific to each sex (deepening of the voice and facial hair growth in boys, and breast growth and menarche in girls). Responses were given on 3- or 4-point scales, ranging from 0 (‘No’) to 3 (‘Yes, definitely’) or 4 (‘Development completed’). Responses to the five questions were averaged. Two female teenagers did not answer whether they had started menstruating, but both indicated at what age they had started menstruating (in both cases, years before participating in the study). We therefore coded their answers on whether they had started menstruating as ‘Yes, definitely’. In both datasets, the pubertal development variable was slightly skewed. In the child dataset, the variable was therefore log 10-transformed, but transformation was not necessary in the teenager dataset.

The 39-question Stress Scale comprises questions about stressful life events, such as death of a parent, parental divorce, and birth of a sibling, where each life event is weighted according to its’ perceived stress. Answers to the items on the Stress Scale were summed up, and higher values denote increased life stress. In the child dataset, the stress variable was skewed and square root-transformed to normality; in the teenager dataset, it was normally distributed.

Rated facial traits. Facial images of parents were cropped and masked along the outline of the face, excluding the hairline and including approximately the top inch of the neck. It should be noted that many fathers were either bald or showed receding hairlines, and the delineation of the hairline was therefore an estimate for these individuals. The images were rated by observers for masculinity (in men) and femininity (in women), attractiveness, perceived health, and physical dominance. Dominance was defined as the probability that the person in the image would win a fistfight with an average person of their sex and age (definition taken from Puts et al., 2006). Each trait was rated by observers on a scale from 1 (e.g., ‘not attractive at all') to 7 (e.g., 'very attractive'). Observers had the option to skip rating faces they recognised, but this never happened. Due to the high number of trials and to avoid rating fatigue, each observer only rated 1/3 of the faces, selected randomly, on each trait, totalling 236 trials/observer (although it should be noted that several observers chose not to complete all trials). Image order was randomised across observers. Order of trait presentation was varied across observers but not in a randomised manner. Male and female faces were shown and rated separately. Each face was rated a minimum of 20 times on each trait.

Observers were recruited online among Durham University Psychology students. Ratings of all parents' faces were completed prior to any teenagers from the university sample commenced their studies in the department, thereby ensuring that no observers were asked to rate the faces of their own or their peers' parents. Observers were not renumerated but had the option to enter a £30 prize draw. In total, 76 observers took part (29 male; age range of all observers 18-30 years, *M* = 22.01, *SD* = 3.39). Eighty percent of the observers were white.

On each facial trait, ratings were excluded for observers who had given the same rating to >80% of the faces. Across 10 traits, ratings from nine observers were excluded. Average ratings were calculated for each face on each trait. Rated facial traits were normally distributed except father's facial dominance, which showed mild skew and was retained without being transformed, and mother's facial attractiveness which was log 10-transformed.

As mentioned previously, a subset of the TVBS parents were photographed both in 2011 and 2018, in which case the best quality image was selected. All parents who had been photographed in 2011 had also been rated for masculinity/femininity, attractiveness, and perceived health in 2012 by another sample of observers. For consistency, we chose to have the entire final sample of parents rated in 2019. Correlations between average ratings in 2012 and 2019 were high: > .5 for all traits (see Table 4.3).

Table 4.3

Pearson's bivariate correlations between 2012 and 2019 ratings of parents' facial traits in the British parents sample (Study 5).

	Masculinity/ femininity	Attractiveness	Health
Fathers	.512**	.710***	.749***
Mothers	.727***	.742***	.597***

** $p < .01$, *** $p < .001$

4.2.1.3. Statistical analysis

Means and standard deviations for the study variables can be viewed in Table 4.4.

Since each observer only rated a subset of all the parents and observer identities therefore differed between parents and facial traits, we attempted to run mixed models controlling for observer identity. However, this was not successful as models failed to converge. Simplified analyses were therefore run by calculating the average rating for each face, on each trait, in a mixed model where parent identity was a fixed effect and observer identity was a random effect. This gave an estimated average rating for each face on each trait, controlling for observer variance.

As mentioned previously, there were two different age controls for parents' traits: offspring age group (child vs teenager) when the parent was photographed for facial traits, and parents' current age for the anthropometric measurements. Similarly to Study 2, in order to control for parental age, we ran linear regression models with parental age as the predictor and each parental trait as the outcome; the saved residuals were then used in all subsequent analyses. Father's facial hair and mother's makeup use were similarly controlled for in all analyses including facial traits.

Table 4.4

Descriptive statistics for all study variables in the British parents sample (Study 5).

Father's traits	*N*	Min	Max	Mean	*SD*
Facial masculinity	68	3.22	5.75	4.59	0.64
Facial attractiveness	68	1.21	4.93	2.63	0.79
Facial health	68	1.73	5.78	4.05	0.85
Facial dominance	68	3.01	5.99	4.24	0.70
Handgrip strength	41	33.00	63.00	47.78	6.60
Height	41	165.50	195.30	177.50	7.40
Mother's traits	*N*	Min	Max	Mean	*SD*
Facial femininity	107	1.44	5.90	3.69	0.96
Facial attractiveness	107	1.46	5.51	3.04	0.86
Facial health	107	1.93	6.02	4.00	0.86
Facial dominance	107	2.40	5.55	4.12	0.63
Handgrip strength	86	18.00	48.00	29.80	5.72
Height	85	149.80	178.30	163.00	6.42
Offspring variables (child dataset)	*N*	Min	Max	Mean	*SD*
Autoimmune illnesses	87	0	7.00	0.77	1.15
Days with a cold in the past 3 months	87	0	20.00	1.72	3.09
Days with the flu in the past 3 months	87	0	10.00	0.32	1.57
Days with gastroenteritis in the past 3 months	87	0	5.00	0.55	1.31
Days on antibiotics in the past 3 months	87	0	10.00	0.69	2.10
Illness composite in the past 3 months	87	0	35.00	3.29	5.16
Bouts of colds and flu in the past 3 years	87	0	12.00	3.07	2.93
Bouts of gastroenteritis in the past 3 years	87	0	12.00	2.28	2.34
Instances of antibiotic use in the past 3 years	87	0	8.00	1.56	1.69
Illness composite in the past 3 years	87	0	26.00	6.91	5.08
Romantic interest	95	0	6.00	1.96	2.05
Control variables (child dataset)	*N*	Min	Max	Mean	*SD*

SES	95	14.00	66.00	37.74	12.70
Stress scale	87	0	1385.00	270.00	273.30
Pubertal development	94	0	1.60	0.55	0.40
Offspring variables (teenager dataset)	*N*	Min	Max	Mean	*SD*
Autoimmune illnesses	96	0	4.00	0.75	0.88
Days with a cold in the past 3 months	96	0	7.00	1.05	1.79
Days with the flu in the past 3 months	96	0	4.00	0.06	0.45
Days with gastroenteritis in the past 3 months	96	0	30.00	0.61	3.40
Days on antibiotics in the past 3 months	96	0	90.00	1.41	9.40
Illness composite in the past 3 months	96	0	93.00	3.14	10.17
Bouts of cold in the past 4 years	95	0	12.00	3.51	2.55
Bouts of flu in the past 4 years	95	0	4.00	0.35	0.85
Bouts of gastroenteritis in the past 4 years	94	0	10.00	1.74	2.10
Instances of antibiotic use in the past 4 years	96	0	4.00	0.84	1.09
Illness composite in the past 4 years	94	0	19.00	6.47	4.28
Prospective bouts of cold	72	0	4.00	0.96	0.98
Prospective bouts of flu	72	0	2.00	0.04	0.26
Prospective bouts of gastroenteritis	71	0	2.00	0.20	0.47
Prospective instances of antibiotic use	71	0	2.00	0.07	0.31
Prospective illness composite	70	0	6.50	1.29	1.38
Romantic interest	78	0	6.00	4.59	1.45
Sexual activity	103	0	9.00	6.39	2.87
Control variables (teenager dataset)	*N*	Min	Max	Mean	*SD*
Stress scale	105	0	659.00	286.58	138.63
Pubertal development	95	1.40	2.80	2.06	0.33

Note. SES = socioeconomic status.

For the TVBS sample's prospective health variables, the amount of time that had passed in between their health being measured in the 2018 lab visit and at the 2019 follow-up was controlled for by saving the residuals of linear regression models with amount of time as the predictor and the respective health variables as outcomes; the saved residuals were then used as outcome measures in the analyses.

To test the hypotheses that dimorphic/attractive parental traits predict better health and earlier sexual onset in offspring, we first ran Pearson's correlations with the saved residuals for each parental trait as predictor and each offspring health and sexual onset variable as the outcome, separately for each offspring age group (child versus teenager). Our pre-registered analysis plan was to proceed by conducting multiple regression models, with the inclusion of control variables, for associations that were significant in simple regression models (e.g., testing associations between father's traits and offspring outcomes, controlling for mother's traits). However, upon checking correlations, we identified only a very limited number of significant associations between parental traits and offspring outcomes; none of which survived computation of *q*-values. We therefore did not proceed with multiple regression analyses.

Lastly, since visual inspection of scatter plots suggested the presence of some nonlinear associations in our data, we performed exploratory curve estimations.

4.2.2. Results

Child dataset. Correlations between parental traits and offspring health and sexual onset – indexed here by romantic interest – showed very little support for the hypotheses (Table 4.5). Across 84 analysed associations, we detected just five significant associations between parental traits and offspring outcomes. (It can be noted that the extremely few significant effects are reflected in a very skewed distribution of *q*-values, meaning that no significant associations remained.) Four of the initially significant correlations comprised father's traits; all of which were in the predicted direction. Father's increased facial dominance predicted significantly fewer autoimmune illnesses and fewer days of the composite measure of common illnesses experienced by offspring in the past three months *(r* = -.316, *p* = .034 and *r* = -.360, *p* = .015, respectively; both *q* = .909). Similarly, fathers with greater handgrip strength had offspring with significantly fewer autoimmune illnesses *(r* = -.526, *p* = .012, *q* = .909), and offspring of taller fathers had experienced fewer days of illness overall in the past three months (*r* = -.441, *p* = .040, *q* = .909). There were three trends whereby more facially healthy-looking fathers had offspring experiencing *more* bouts of colds and flu and more illnesses overall in the past three years (*r* = .293, *p* = .051, *q* = .909; *r* = .253, *p* = .093, *q* = .971, respectively), and whereby offspring of stronger fathers had used antibiotics less frequently in the past three years (*r* = -.391, *p* = .072, *q* = .911). Offspring of mothers with greater handgrip strength reported lower romantic interest (*r* = -.295, *p* = .027, *q* = .909). There were also trends for more facially dominant mothers to have offspring who used more antibiotics and reported lower romantic interest (*r* = .236, *p* = .052, *q* = .909; *r* = .214, *p* = .070, *q* = .910) and for the offspring of stronger mothers to take antibiotics more frequently (*r* = .265, *p* = .061, *q* = .909) For the control variables – socioeconomic status, offspring stress, and offspring pubertal development – a single significant association was observed: offspring from families with higher socioeconomic status had used antibiotics on fewer occasions in the past three years *(r* = -.251, *p* = .019, *q* = .909).

Overall, while nonsignificant associations should be interpreted with caution, the direction of the linear associations was generally inconsistently positive and negative across parental traits and offspring outcomes. The exceptions to this were father's rated facial dominance and father's handgrip strength, which were consistently negatively linked to offspring illness. Notably, in contrast to predictions based on the ICHH, father's facial masculinity did not show consistent associations with offspring health.

Table 4.5

Pearson's bivariate correlations between parental traits, control variables, and offspring variables in the British child dataset, controlling (where relevant) for parental age, father's facial hair, and mother's makeup use (Study 5).

Father's traits	Auto-immune illnesses	Illness comp.: 3 months	Colds/ flu: 3 years	Gastro-enteritis: 3 years	Anti-biotics: 3 years	Illness comp.: 3 years	Romantic interest
Facial masculinity	-.144 $N = 45$	.052 $N = 45$	-.001 $N = 45$	.199 $N = 45$	.180 $N = 45$	.157 $N = 45$	.007 $N = 50$
Facial attractiveness	.061 $N = 45$	.232 $N = 45$	.245 $N = 45$	.176 $N = 45$	.076 $N = 45$	.189 $N = 45$	.112 $N = 50$
Facial health	-.080 $N = 45$	.253 $N = 45$	.293 $N = 45$	.124 $N = 45$	-.054 $N = 45$	.178 $N = 45$	.062 $N = 50$
Facial dominance	-.316* $N = 45$	-.360* $N = 45$	-.063 $N = 45$	-.054 $N = 45$	-.063 $N = 45$	-.096 $N = 45$	-.081 $N = 50$
Handgrip strength	-.526* $N = 22$	-.359 $N = 22$	-.058 $N = 22$	-.353 $N = 22$	-.391 $N = 22$	-.254 $N = 22$	.032 $N = 25$
Height	-.162 $N = 22$	-.441* $N = 22$	-.067 $N = 22$	.038 $N = 22$	-.324 $N = 22$	-.094 $N = 22$	.056 $N = 25$

Mother's traits	Auto-immune illnesses	Illness comp.: 3 months	Colds/ flu: 3 years	Gastro-enteritis: 3 years	Anti-biotics: 3 years	Illness comp.: 3 years	Romantic interest
Facial femininity	.027 $N = 68$	.084 $N = 68$	.009 $N = 68$	.052 $N = 68$	-.115 $N = 68$	-.017 $N = 68$	.000 $N = 73$
Facial attractiveness	.081 $N = 68$	.054 $N = 68$	-.035 $N = 68$	.090 $N = 68$	-.120 $N = 68$	-.019 $N = 68$	-.006 $N = 73$
Facial health	.093 $N = 68$	.088 $N = 68$	-.027 $N = 68$	.071 $N = 68$	-.148 $N = 68$	-.012 $N = 68$	.027 $N = 73$
Facial dominance	-.181 $N = 68$	-.148 $N = 68$	.082 $N = 68$	-.082 $N = 68$	.236 $N = 68$	.085 $N = 68$	-.214 $N = 73$
Handgrip strength	-.042 $N = 51$	.229 $N = 51$	-.062 $N = 51$	.059 $N = 51$	.265 $N = 51$	.058 $N = 51$	-.295* $N = 56$
Height	-.204 $N = 50$	-.067 $N = 50$	-.120 $N = 50$	-.030 $N = 50$	.103 $N = 50$	-.072 $N = 50$	-.006 $N = 55$

Control variables	Auto-immune illnesses	Illness comp.: 3 months	Colds/ flu: 3 years	Gastro-enteritis: 3 years	Anti-biotics: 3 years	Illness comp.: 3 years	Romantic interest
SES	.103 *N* = 87	.025 *N* = 87	.012 *N* = 87	.035 *N* = 87	-.251* *N* = 87	-.080 *N* = 87	-.196 *N* = 95
Stress scale	.031 *N* = 87	.053 *N* = 87	.150 *N* = 87	.106 *N* = 87	.052 *N* = 87	.163 *N* = 87	.065 *N* = 87
Pubertal development	-.158 *N* = 86	.015 *N* = 86	.034 *N* = 86	-.074 *N* = 86	-.046 *N* = 86	-.040 *N* = 86	.109 *N* = 94

Note. Comp = composite; HGS = handgrip strength; masc = masculinity; SES = socioeconomic status. * $p < .05$. No associations remained significant after computation of *q*-values.

We detected three significant quadratic associations, none of which remained significant after corrections. There was a significant threshold effect for father's handgrip strength on offspring's autoimmune illnesses, so that the strongest fathers had the healthiest offspring [$F(2,19) = 3.653$, $p = .045$, $R^2 = .278$]. Similarly, mothers whose facial dominance was rated as intermediate had offspring with more autoimmune illnesses [$F(2,65) = 3.317$, $p = .042$, $R^2 = .093$] and lower romantic interest [$F(2,70) = 4.138$, $p = .020$, $R^2 = .106$].

Teenager dataset. Similarly to the child dataset, 140 simple regression models revealed just six significant associations between parental traits and offspring outcomes, controlling for parental age (where possible), father's facial hair, and mother's makeup use. In line with predictions, offspring of more facially dominant-looking fathers had experienced fewer days with cold symptoms in the past three months ($r = -.310$, $p = .021$, $q = .909$). Fathers with greater handgrip strength had offspring with fewer prospective colds and fewer prospective bouts of common illnesses overall ($r = -.402$, $p = .038$, $q = .909$ for both associations). Offspring of facially attractive and healthy-looking fathers also reported greater sexual activity ($r = .317$, $p = .018$, $q = .909$ and $r = .295$, $p = .029$, $q = .909$, respectively). The only significant effect observed for a maternal trait was in the opposite direction: mothers with greater handgrip strength had offspring who had experienced more bouts of gastroenteritis in the past four years (r = .233, p = .034, q = .909). Again, none of these effects remained significant after computation of q-values, and we therefore did not run multiple regression models. The directions of the associations between parental traits and offspring outcomes were, again, inconsistent.

The control variable life stress was positively and significantly associated with offspring's greater romantic interest and greater sexual activity (r = .288, p = .011, q = .909 and r = .513, p < .001, q < .001, respectively). Across the entire Study 5 analyses, the latter effect was the only one that survived corrections for multiple comparisons.

Lastly, we observed five significant quadratic associations, none of which remained after computation of q-values. Offspring of fathers with intermediate handgrip strength had experienced fewer bouts of gastroenteritis in the past four years [F(2,36) = 4.982, p = .012, R^2 = .217]. Moderately tall fathers also had offspring who had experienced fewer days of illness in the past three months [F(2,37) = 3.989, p = .027, R^2 = .177]. There were threshold effects for two maternal traits: offspring of mothers with the greatest handgrip strength had experienced more bouts of gastroenteritis in the past four years [F(2,80) = 3.217, p = .045, R^2 = .074], and the most facially attractive mothers had offspring who had had lower prevalence of common illnesses in the past three months [F(2,82) = 3.208, p = .046, R^2 = .073]. Mothers with intermediate facial health had offspring who had experienced the least illness in the past three months [F(2,82) = 4.320, p = .016, R^2 = .095].

Table 4.6

Pearson's bivariate correlations between parental traits, control variables, and offspring variables in the British teenager dataset, controlling (where relevant) for parental age, father's facial hair, and mother's makeup use (Study 5).

Father's traits	Auto-immune illnesses	Illness comp.: 3 mos	Colds: 4 years	Gastro-enteritis: 4 years	Anti-biotics: 4 years	Illness comp.: 4 years	Colds: prosp.	Illness comp.: prosp.	Rom. interest	Sexual activity
Facial masculinity	-.001 *N* = 55	-.135 *N* = 55	.130 *N* = 54	-.016 *N* = 54	-.011 *N* = 55	.012 *N* = 54	.060 *N* = 42	.029 *N* = 42	-.282 *N* = 42	-.108 *N* = 55
Facial attractiveness	-.081 *N* = 55	-.080 *N* = 55	.055 *N* = 54	.105 *N* = 54	-.115 *N* = 55	.099 *N* = 54	.038 *N* = 42	-.007 *N* = 42	.076 *N* = 42	.317* *N* = 55
Facial health	-.091 *N* = 55	-.114 *N* = 55	.098 *N* = 54	-.037 *N* = 54	-.084 *N* = 55	.072 *N* = 54	-.027 *N* = 42	-.069 *N* = 42	.067 *N* = 42	.295* *N* = 55
Facial dominance	-.157 *N* = 55	-.310* *N* = 55	.095 *N* = 54	-.113 *N* = 54	-.082 *N* = 55	-.049 *N* = 54	-.043 *N* = 42	-.113 *N* = 42	-.243 *N* = 42	.071 *N* = 55
Handgrip strength	.005 *N* = 40	.033 *N* = 40	.253 *N* = 39	-.210 *N* = 39	-.125 *N* = 40	-.016 *N* = 39	-.402* *N* = 27	-.402* *N* = 27	.074 *N* = 27	.188 *N* = 40
Height	.100 *N* = 40	.026 *N* = 40	-.136 *N* = 39	-.134 *N* = 39	-.115 *N* = 40	-.211 *N* = 39	.055 *N* = 27	.066 *N* = 27	-.045 *N* = 27	-.086 *N* = 40

Mother's traits	Auto-immune illnesses	Illness comp.: 3 mos	Colds: 4 years	Gastro-enteritis: 4 years	Anti-biotics: 4 years	Illness comp.: 4 years	Colds: prosp.	Illness comp.: prosp.	Rom. interest	Sexual activity
Facial femininity	.015 *N* = 85	-.116 *N* = 85	.112 *N* = 85	-.177 *N* = 84	-.122 *N* = 85	-.085 *N* = 84	-.119 *N* = 69	-.126 *N* = 67	.062 *N* = 74	.100 *N* = 85
Facial attractiveness	-.037 *N* = 85	-.137 *N* = 85	.102 *N* = 85	-.134 *N* = 84	-.123 *N* = 85	-.065 *N* = 84	-.101 *N* = 69	-.126 *N* = 67	-.036 *N* = 74	.080 *N* = 85
Facial health	.000 *N* = 85	-.146 *N* = 85	.092 *N* = 85	-.147 *N* = 84	-.060 *N* = 85	-.071 *N* = 84	-.093 *N* = 69	-.085 *N* = 67	-.065 *N* = 74	.009 *N* = 85
Facial dominance	-.135 *N* = 85	.039 *N* = 85	-.183 *N* = 85	.018 *N* = 84	-.032 *N* = 85	-.120 *N* = 84	.075 *N* = 69	.079 *N* = 67	-.058 *N* = 74	-.080 *N* = 85
Handgrip strength	-.076 *N* = 84	.103 *N* = 84	.060 *N* = 84	.233* *N* = 83	.008 *N* = 84	.162 *N* = 83	.051 *N* = 68	.055 *N* = 66	.179 *N* = 73	.053 *N* = 84
Height	-.019 *N* = 83	.023 *N* = 83	.057 *N* = 83	.077 *N* = 82	.136 *N* = 83	.153 *N* = 82	.101 *N* = 67	.139 *N* = 65	.056 *N* = 72	.023 *N* = 83

Control variables	Auto-immune illnesses	Illness comp.: 3 mos	Colds: 4 years	Gastro-enteritis: 4 years	Anti-biotics: 4 years	Illness comp.: 4 years	Colds: prosp.	Illness comp.: prosp.	Rom. interest	Sexual activity
Stress scale	.032 *N* = 96	-.123 *N* = 96	-.183 *N* = 95	.001 *N* = 94	.108 *N* = 96	-.094 *N* = 94	-.159 *N* = 72	-.129 *N* = 70	.288* *N* = 78	**.513***** ***N* = 103**

Pubertal dev.	.123 *N* = 88	.056 *N* = 88	-.177 *N* = 87	.080 *N* = 86	.006 *N* = 88	-.102 *N* = 86	.101 *N* = 66	.102 *N* = 64	.017 *N* = 70	.128 *N* = 93

Note. Comp = composite; dev = development; masc = masculinity; mos = months; prosp = prospective; rom = romantic.
* $p < .05$, ** $p < .01$, *** $p < .001$. Associations that remained significant after computation of q-values are bolded.

4.2.3. Interim discussion

In Study 5, we tested the hypothesis that more masculine fathers would have healthier offspring – as indexed by common illnesses – in a longitudinal sample of British families. Since maternal traits might mediate the relationship between paternal traits and offspring health, we also tested for effects of maternal traits. Overall, our results did not support these predictions. Across more than 200 tested associations between parental traits and offspring health outcomes, we initially found just eight that were significant. We observed seven associations where fathers who looked more facially dominant, who were taller, and who had greater handgrip strength had healthier offspring; and one association where mother’s greater handgrip strength predicted worse offspring health. Based on our previous finding that masculine men showed greater mating success (for which earlier sexual onset is a proxy), we also tested for an association between parental traits and timing of offspring sexual onset. We found two associations whereby more attractive and healthy-looking fathers had offspring who were more sexually active, and one whereby stronger mothers had offspring reporting less interest in romantic relationships. None of these effects survived corrections for multiple comparisons, however. The only relationship in the entire analysis to survive multiple-comparison corrections was greater offspring stress predicting greater sexual activity. Thus, after corrections, no parental traits significantly predicted any offspring outcomes. It can be noted that, of the (remarkably few) linear associations between paternal traits and offspring health that were significant prior to correction for multiple comparisons, the majority *were* in the predicted direction – but only three of those associations comprised father’s facial traits (which, as noted above, typically have received more attention in the literature); all of which were for facial dominance. Facial masculinity did not significantly predict any offspring outcomes in the entire analyses, and mother's facial traits similarly showed no significant associations. While we observed a very limited number of quadratic associations, they were not in a consistent direction and none of them survived corrections.

It should be reiterated that sample sizes for some of these analyses were small. If there was a robust relationship between paternal traits and offspring quality, however, we would still expect the direction of the associations to be consistent, even in the absence of power to detect weak, significant effects. Here, we note that the overall direction of the observed effects was not consistent across traits, except possibly for father's perceived facial dominance and handgrip strength, which showed almost exclusively negative correlations with prevalence of offspring illness.

4.3. Study 6: Turkish families

4.3.1. Methods

4.3.1.1. Participants

Participants were the same Turkish families as described in Study 3.

4.3.1.2. Materials and procedure

See Chapter 3 for description of parents' facial traits.

Offspring health variables. Parents reported how many bouts of colds and flu each of their offspring had experienced, as well as instances of serious illness before each offspring was 5 years of age. These variables were then averaged across all offspring for each family. Both variables were skewed; square root-transformations improved but did not eliminate skew.

Control variable. Parents also reported their monthly household income; this variable was normally distributed.

Table 4.7
Descriptive statistics for relevant study variables in the Turkish sample (Study 6).

Offspring health	*N*	Min	Max	Mean	*SD*
Average bouts of colds and flu	26	0	5.33	1.15	1.54
Average instances of serious illness before age 5	26	0	5.00	0.78	1.15
Control variable	***N***	**Min**	**Max**	**Mean**	***SD***
Household income (Turkish lira)	26	11,000	120,000	52,000	27,054.02

4.3.1.3. Statistical analysis

Means and standard deviations for relevant study variables are displayed in Table 4.7. We ran correlations to test the hypotheses that parental dimorphism and attractiveness would predict better offspring health, controlling for parental age, father's facial hair, and mother's makeup use. Lastly, we checked for quadratic relationships.

4.3.2. Results

The results did not support the hypotheses; no parental traits predicted offspring health (Table 4.8). The only significant association was between higher household income and fewer colds and flu experienced by offspring (r = -.402, p = .042, q = .909), but it did not survive corrections for multiple comparisons. There were no significant quadratic associations.

4.3.3. Interim discussion

In Study 6, we thus tested whether parental dimorphic and attractive facial traits were associated with better offspring health, indexed by bouts of colds and the flu and by serious illnesses in early childhood. We did not find evidence that this was the case. It should be noted that sample sizes here were very small. To increase power, we therefore meta-analysed the effects from Study 5 and Study 6.

Table 4.8

Pearson's bivariate correlations between parental traits, control variables, and offspring variables in the Turkish sample, controlling (where relevant) for parental age, father's facial hair, and mother's makeup use (Study 6).

Father's traits	Offspring colds/flu	Offspring serious illness
Facial masculinity	.086 *N* = 26	.256 *N* = 26
Facial attractiveness	-.040 *N* = 26	-.024 *N* = 26
Facial health	-.033 *N* = 26	-.264 *N* = 26
Mother's traits	**Offspring colds/flu**	**Offspring serious illness**
Facial femininity	-.143 *N* = 25	.177 *N* = 25
Facial attractiveness	-.292 *N* = 25	.075 *N* = 25
Facial health	-.168 *N* = 25	.059 *N* = 25
Control variable	**Offspring colds/flu**	**Offspring serious illness**
Household income	-.402* *N* = 26	.145 *N* = 26

Note. * $p < .05$. No associations remained significant after computation of q-values.

4.4. Meta-analytic associations

4.4.1. Statistical analysis

The meta-analysis was set up in an identical way to previous meta-analyses (Chapters 2 and 3). Parental facial dimorphism, attractiveness, and health had been measured in both Study 5 and Study 6 and could therefore be included as predictors. For offspring health variables, we included autoimmune illnesses and composite health measures from Study 5, and both offspring health variables from Study 6 (colds/flu and serious illnesses). In total, this gave us 54 effects, or 9 effects per parental trait. All predicted relationships between parental traits and offspring health were negative. Effects were clustered by sample (British or Turkish).

4.4.2. Results

As can be seen in Table 4.9, we detected no significant relationships between parental traits and offspring health (funnel plots can be viewed in Appendix O). Effect sizes were very weak and centred around zero for all traits.

Table 4.9

Parental traits predicting offspring health across Studies 5 and 6. Pearson's r (95% CI); p value for meta-analytic effect, q-value (correcting for multiple comparisons); number of observations (k), and 'unique' participants (n; but note that some British participants were sampled more than once); test for heterogeneity (Q), p value for heterogeneity. No associations were significant.

	Father's traits		
	Facial masculinity	Facial attractiveness	Facial health
Offspring health	r = .032 (-0.116, 0.180), p = .673, q = .985 k = 9, n = 126 Q(df = 8) = 4.718, p = .787	r = .042 (-0.060, 0.145), p = .419, q = .985 k = 9, n = 126 Q(df = 8) = 4.618, p = .798	r = -.012 (-0.137, 0.112), p = .847, q = .985 k = 9, n = 126 Q(df = 8) = 7.704, p = .463
	Mother's traits		
	Facial femininity	Facial attractiveness	Facial health
Offspring health	r = -.030 (-0.114, 0.054), p = .482, q = .985 k = 9, n = 178 Q(df = 8) = 3.941, p = .862	r = -.046 (-0.130, 0.038), p = .283, q = .985 k = 9, n = 178 Q(df = 8) = 4.645, p = .795	r = -.027 (-0.111, 0.057), p = .530, q = .985 k = 9, n = 178 Q(df = 8) = 4.044, p = .853

Note. k = number of observations; n = number of unique participants; Q = Cochran's Q test of heterogeneity; q = q-value.

4.5. Discussion

The ICHH predicts that highly masculine men should not only be healthier themselves, but they should also pass on their immunocompetence to their offspring who should thus be healthier and more viable. If masculine men are of higher quality, they should also be able to attract better quality partners, which may further mediate the relationship between paternal traits and offspring quality. In this chapter, we reported the results of two studies testing offspring health as a function of parental dimorphic and attractive traits in one British and one Turkish sample. In industrialised populations where offspring mortality is extremely low, health is a more appropriate measure of offspring viability than mortality. Here, we used health measures that have previously been found be associated with men's facial masculinity (Boothroyd et al., 2013; Thornhill & Gangestad, 2006). Therefore, if facial masculinity does reliably reflect heritable immunity, this should translate into healthier offspring for more facially masculine fathers. While a limited number of paternal traits (but notably, not paternal facial masculinity) were significantly associated with better offspring health in the British sample, no associations remained significant after correcting for multiple comparisons. No associations were significant in the Turkish sample, and meta-analysed effects across the two samples showed that effect sizes centred around zero for both paternal and maternal facial traits. Across these two samples, offspring were sampled during the first five years of life, in late childhood, and as teenagers (when offspring's own hormone levels are high): associations between parental traits and offspring health did not appear to differ between these different developmental stages. Thus, we found no support for our predictions here. This suggests that to the extent that facial masculinity does show a relationship with own health (a link which cannot be considered to be conclusively demonstrated at this point), it does not appear to reflect heritable immunity; nor do our results suggest that maternal traits show any relationship with offspring health either. Lastly, since our meta-analysis in Chapter 2 showed that some indices of men's masculinity predict greater mating success, we also tested offspring sexual onset as a function of parental traits here; offspring of facially attractive and healthy-looking fathers reported greater sexual activity, but these associations did not survive corrections, either.

If masculine traits do reliably signal heritable immunocompetence, it follows that women should be attracted to such traits in order to produce better quality offspring, thereby increasing their own reproductive success. While beyond the scope of this thesis, it can be noted that evidence for women actually preferring masculine traits in men is equivocal (e.g., Scott et al., 2010). More crucially in the context of this chapter, our findings fly in the face of the very foundation of this (very commonly cited) prediction: we find no evidence that more masculine men have better quality offspring in the first place. It is important to highlight that both of our samples came from industrialised populations, however. If the results should hold true across populations, this would suggest that to the extent that women *do* prefer more masculine men – and our meta-analytic results in Chapter 2 suggest that *some* indices of men's masculinity *do* increase men's reproductive output – masculinity confers benefits other than heritable immunocompetence. On a related note, in Chapter 3 we observed no relationship between paternal attractiveness and reproductive outcomes; here we find that father's facial attractiveness does not predict offspring quality either. If future research shows that the lack of a relationship between men's attractiveness and reproductive outcomes is robust across populations, this is very problematic for the claim that women's attractiveness preferences have evolved to secure the acquisition of a high-quality mate (or at the very least, begs the question of which mate quality measure attractiveness *does* communicate).

We found just two significant correlations for maternal traits, whereby stronger mothers had offspring in worse health and with less romantic interest; like other significant correlations here, it did not survive corrections. To reiterate, facial femininity is strongly attractive in female faces and commonly claimed to signal good quality, but maternal facial traits do not predict any aspect of better quality in offspring here. Moreover, our review in Chapter 3 did not show any consistent support for the notion that facial traits in women communicate health. Taken together, this appears to suggest that whatever quality facial femininity/attractiveness in women does signal, it has very little to do with health. Similarly to the point made in the previous paragraph with regards to women's preferences for men's attractiveness, it seems unlikely that men's preferences for feminine/attractive facial traits in women is an adaptation to secure healthier mates and thereby producing healthier offspring. With regards to women's feminine body traits, evidence is so far lacking.

Lastly, we note that the associations with offspring health for father's rated facial dominance and father's handgrip strength appeared to be more consistently linked to better offspring health compared to the other traits we tested. It is possible that we lacked the power to detect significant effects for these relationships here, and these associations should therefore be explored in future research.

Chapter 5.

Male morphology and attractiveness as indicators of partner quality

---

5.1. Introduction

5.1.1. Predictions based on extant literature

In the previous chapters, we tested different pathways through which more masculine men may achieve greater reproductive success. While we did not find evidence suggesting that more masculine men or more feminine women have better quality offspring in Chapter 4, Chapter 2 did show that men with more masculine bodies have increased reproductive success as well as greater partner numbers. As initially raised in Chapter 3, it has been suggested that the greater reproductive output by more masculine men might be mediated by these men having access to more fertile mates (Rosenfield et al., 2020). This could thus be a complementary route for men to achieve greater biological fitness, in addition to greater mating success.

Why would we expect more masculine men to be able to acquire more fertile mates? To reiterate, extant hypotheses proposed to explain the evolution/maintenance of masculine traits in human males posit that such traits signal or cue mate quality. If more masculine men are indeed of higher quality (regardless of whether this is through superior immunocompetence or through increased same-sex competitiveness), that should afford more masculine men not only increased mating success, but also greater choice when selecting a partner. Thus, more masculine men should be able to acquire higher-quality partners (as mentioned previously, oftentimes presumed to correlate with fertility), resulting in assortative mating for indicators of mate quality.

5.1.2. Evidence for assortative mating on morphological traits in humans

Assortative mating refers to non-random mating, whereby the two individuals in a mating dyad display traits that correlate with other more than can be expected by chance (Thiessen & Gregg, 1980). Such assortment may be positive or negative, whereby positive assortative mating (also referred to as ‘homogamy’) means that the two individuals display similar phenotypic traits. Conversely, negative (or disassortative) mating – ‘heterogamy’ − means that two mating individuals show negatively correlated traits. Underlying causal mechanisms for how and why assortative mating occurs generally are beyond the scope of this thesis and will not be reviewed here (but are oftentimes argued to have evolved due to increasing biological fitness, although other explanations are also possible: see e.g., Silventoinen et al., 2003; Štěrbová & Valentova, 2012). Here, we focus on the question of whether humans show assortative mating for morphological indices of mate quality.

Assortative mating for morphological traits in humans has been demonstrated in terms of, for example, height (Robinson et al., 2017; Stulp et al., 2016), body weight/body mass index (BMI: Fisher et al., 2014; Robinson et al., 2017; Silventoinen et al., 2003), eye and hair colour (Little et al., 2003; but see Štěrbová et al., 2019, for evidence suggesting that this may be due to sexual imprinting on parental traits rather than assortment on own traits), and physical attractiveness (Feingold, 1988), with significant correlations ranging between $r = .10$ to .50. All of these traits have been found to show positive assortment, whereby the two individuals in a couple resemble one another.

Based on extant evidence for positive assortment on morphological traits, it may be expected that humans should also assort on sexually dimorphic traits. Making theoretically derived predictions for the direction of such assortment is more challenging, however: should we expect a sexually dimorphic species, like humans, to assort on the degree of sexual dimorphism (i.e., that more masculine men are partnered with more feminine women) or to assort on masculinity/femininity separately (i.e., that more masculine men are partnered with more masculine women, and vice versa)? To date, there is evidence for both predictions, although it should be noted that this is not a well-researched area. Across two studies, Cornwell and Perrett (2008) showed that middle-aged married U.K. couples assorted on facial attractiveness, judging from non-standardised, participant-provided photographic stimuli. The authors also observed a nonsignificant trend for assortment on rated facial dimorphism (so that more facially masculine men were partnered with more facially feminine women) in one of their studies, but not in the other. In contrast, Burriss and colleagues (2011) tested whether more facially feminine and symmetric women were more likely to have more masculine and symmetric male partners (i.e., not assort on facial femininity/masculinity per se, but on sexual dimorphism as well as symmetry). In two samples comprising 34 U.K. and 110 U.S. young heterosexual couples, they observed that couples showed significant positive assortment for facial symmetry but not masculinity (but it can be noted that the correlation between the two partners' facial masculinity was positive in both samples, indicated possible positive assortment for masculinity rather than dimorphism). In a similar vein, Little and colleagues (2006) reported that in a sample of 85 married couples of varying ages in the U.K., they found assortment on rated facial masculinity, but this association did not remain significant when controlling for facial attractiveness.

In terms of assortment for body dimorphism (aside from the well-established assortment for height, as mentioned previously), we are only aware of a handful of studies testing this. Couples may show some assortment for 2D:4D ratios (Richards et al., 2020; Voracek et al., 2007); in addition, one recent study showed that men with more masculine

2D:4D ratios had female partners with a significantly lower WHR (*N* = 50 young couples; Kuna & Galbarczyk, 2018). In the latter study, men with a more masculine right-hand 2D:4D ratio were also significantly more likely to have a partner with larger breasts and a narrow waist, suggesting assortment on sexual dimorphism across traits. Thus, the evidence so far appears to support either prediction: that more masculine men may have more masculine *or* more feminine female partners.

How might such assortment work in practice? If e.g., sexual dimorphism/attractiveness communicates mate quality, it could be theorised that while all individuals might *prefer* a high quality (e.g., highly dimorphic/attractive) mate, assortment on indices of mate quality means that not all individuals will be able to actualise their preferences. Rather, individuals will have no choice but to couple up with the best quality mate that is available to them: they will effectively end up assorting on quality, regardless of what their true preferences are − and individuals with higher mate value will thus be more able to realise their preferences (e.g., Gangestad & Simpson, 2000). Alternatively, it is also plausible that people's mate preferences may be modulated by their own traits (possibly as a calibration of preferences based on own quality: Bailey et al., 2011), which should in turn influence actual mate choices. This could, in turn, modulate assortment for dimorphic traits. For example, women's own voice pitch might influence the degree to which they prefer masculinity in a man's voice pitch (Vukovic et al., 2010), and both men and women who perceive themselves to be more attractive also express preferences for increased sexual dimorphism in opposite-sex faces (Marcinkowska et al., 2021). It should be noted here, however, that mate *preferences* may not always be reflected in actual mate *choices* (Courtiol et al., 2010; Stulp et al., 2013), and it is therefore necessary to also directly measure phenotypic associations between actual couples.

5.1.3. The present chapter

To reiterate, one pathway for masculine traits to be selected for is if more masculine men have access to better quality partners. Assortative mating for a range of traits, such as attractiveness and height, has been demonstrated in humans, but assortative mating for sexually dimorphic traits is less commonly studied. Theoretically driven predictions could be made both for assortment on sexual dimorphism (so that more masculine men are partnered with more feminine women) as well as separately for masculinity vs femininity (so that more masculine men are partnered with more masculine women and vice versa). The limited available evidence does not show clear support for either prediction. Furthermore, extant literature shows a lack of standardised methods, making it challenging to assess how accurate the results are or compare across studies.

Here, we tested the prediction that couples assort on quality measures across two U.K. samples: namely, that *i.* men and women mate assortatively on dimorphism/attractiveness, *ii.* more masculine/attractive men have better quality partners, indexed by increased femininity/attractiveness, higher mate value, and lower incidence of common illnesses, and *iii.* more dimorphic/attractive men and women or of higher quality themselves, using the same quality indices. Mate quality was indexed by facial and anthropometric traits commonly argued to correlate with either immunocompetence, same-sex competitiveness, and/or fertility (facial traits: attractiveness, dimorphism, perceived health, dominance, and prestige; handgrip strength, height, BMI, shoulder circumference [in men] and WHR [in women]) as well as self-reported incidence of common illnesses and self- and partner-rated mate value. In Study 7, we tested these predictions in a sample of U.K. student couples ($N = 100$); Study 8 sampled the British parents reported in Chapter 4.

5.2. Study 7: Student couples

5.2.1. Methods

5.2.1.1. Participants

The sample consisted of 100 romantic, heterosexual couples (total *N* = 200 participants; male age range 18-29, *M* = 20.73, *SD* = 2.04; female age range 18.27, *M* = 20.36, *SD* = 1.71). The couples were recruited among the Durham University student population, and at least one individual in each couple was a student at Durham University. Eighty-one percent of the total sample was white. Couples had to have been in a relationship lasting a minimum of four weeks to be eligible to take part (but see information about relationship duration below). Male and female participants stated how they defined their relationship with their partner and the duration of the relationship separately (Table 5.1). Ninety-six men defined their relationship as ‘committed’ and two men that they were ‘living together as married’ with their partner, whereas 95 women stated that they considered their relationship with their partner as ‘committed’, three that they were ‘casually dating’, and two that they were ‘living together as married’. Relationship duration varied from under one month to more than three years. Most couples stated that they had been together between seven and 12 months (40 men and 43 women) or one to two years (24 men and 23 women).

Table 5.1

Relationship definition and duration as stated by men and women from the student couples (Study 7).

Relationship status	Men	Women
Casually dating	0	3
Committed relationship	96	95
Living together as married	2	2
Total	98	100

Relationship duration	Men	Women
Less than 1 month	1	1
1-3 months	7	9
4-6 months	7	6
7-12 months	40	43
1-2 years	24	23
2-3 years	12	9
More than 3 years	5	6
Not sure	2	2
Prefer not to say	0	1
Total	98	100

5.2.1.2. Materials and procedure

The study protocol was approved by Durham University Department of Psychology Ethics Committee. All participating couples gave informed consent and were paid £20/couple for participation. Data collection took place in the lab during one session which took approximately one hour. The two individuals in each couple participated in the session together, but they were separated for all parts of the study. Several elements of the study protocol were not relevant to this study and are not reported here. Participants were free to omit any parts of the study protocol they did not want to do. All self-report measures were completed in a private room, without the experimenter present. One participant's self-report measures were omitted from analysis due to language issues. See descriptive statistics for all relevant study variables in Table 5.2.

Facial images. Facial photographs were taken of participants; the methods were identical to the methods used in the British parents sample.

Rated facial traits. Facial images were rated by observers using the same methods as in the British parents sample. In addition to facial masculinity/femininity, attractiveness, perceived health, and physical dominance, this sample was also rated for perceived facial prestige, where observers were informed that 'a prestigious person tells other people what to do, is respected, influential, and often a leader; while non-prestigious people are not influential or assertive and are usually directed by others' (adapted from Mazur et al., 1994). A total of 102 observers rated the faces; after excluding four observers who gave the same rating to >80% of all the faces they rated across all traits, or who only used the extreme end of response scales throughout the study, the total sample consisted of 98 observers (42 male; age range of all observers 17-31 years, *M* = 22.15, *SD* = 3.76). Sixty-seven percent of the observers were white. One observer completed the study twice; all of their ratings were retained. The study comprised 250 trials/observer, but not all observers completed all trials. There were also signs of rating fatigue. Therefore, on each individual facial trait, ratings from observers who had given the same rating on >80% trials, or who had only used extreme ends of the response scales (i.e., 1-3 or 5-7) for all faces on a given trait were excluded. On 51 trials (0.2% of all trials), the observer recognised the individual in the image and skipped rating the face. After data cleaning, the faces had been rated by a minimum of nine observers/trait (mean number of observers rating each face on each trait = 19). Average ratings for each face on each trait was calculated in the same manner as for the British parents sample. All facial trait ratings for both sexes were normally distributed.

Anthropometric measurements. All participants' handgrip strength and height were measured, using the same methods as in the British parents sample. In men, shoulder circumference was measured at the widest part over the shoulder blades with the participant's arms relaxed to the sides; the measurement was taken twice and averaged. In women, measurements of the waist – taken twice at the narrowest part between the ribcage and the navel – and hips were taken twice, averaged, and WHR subsequently calculated. The same female experimenter took the measurements for the whole sample. Remeasurement reliability was high (male shoulders: r = .987; female waist: r = .992; female hips: r = .993). Participants' self-reported weight was used to calculate BMI. Weight data that was clearly incorrect was omitted for three participants (weight given as below 5 kg, or weight that would give the participant a BMI that was very severely underweight when it was clearly not the case). When weight was given by the participant as a range (e.g., 68-70 kg), the midpoint was used.

HGS in both sexes, men's shoulder circumference, and men's height were normally distributed. WHR, women's height, and BMI in both sexes were skewed and log 10-transformed.

Self-perceived mate value. Participants completed the Self Perceived Mating Success Scale (Landolt et al., 1995; Appendix P) about both themselves and about their partner, resulting in self- and partner-rated mate value for both sexes. The scale in question contains eight items intended to measure one's self-perceived mate value, such as 'Members of the opposite sex that I like, tend to like me back' and 'I can have as many sexual partners as I choose'. Participants responded on a 5-point scale, from 'Strongly disagree' to 'Strongly agree'. All mate value measures were normally distributed except men's self-rated mate value, which was mildly skewed and retained without being transformed.

Health variables. Participants self-reported how many bouts of colds, flu, and gastroenteritis they had experienced in the past three months as well as how many times they had been prescribed antibiotics in that time. Responses were multiple-choice; for the option ‘5 or more’, this was recoded as ‘5’. If more than one response option was chosen (this was possible on one question due to an error in the survey), the two options were averaged (e.g., participant had chosen both ‘0’ and ‘1’, a value of ‘0.5’ was assigned). Like in the British family sample, a composite measure of all illness measures was created. All health variables were significantly skewed; colds, gastroenteritis, and illness composite variables were square root- or log 10-transformed to approach normality, whereas flu and antibiotics could not be transformed.

5.2.1.3. Statistical analysis

Means and standard deviations for the study variables can be viewed in Table 5.2.

Table 5.2

Descriptive statistics for all study variables in the student couples (Study 7).

Men’s variables	*N*	Min	Max	Mean	*SD*
Facial masculinity	99	2.56	5.62	4.11	0.73
Facial attractiveness	99	1.78	4.96	3.05	0.66
Facial health	99	2.68	5.65	4.28	0.71
Facial dominance	99	2.07	5.32	3.61	0.74
Facial prestige	99	2.20	5.02	3.35	0.61
Handgrip strength	100	23.50	65.00	45.78	7.77
Height	100	160.40	196.00	180.58	6.26
Weight	93	54.50	117.00	75.61	11.54
BMI	92	17.92	35.80	23.08	3.32
Shoulder circumference	99	103.45	131.50	116.08	6.46
Bouts of colds in the past 3 months	98	0	5	1.44	1.08
Self-rated mate value	97	1.00	4.63	2.98	0.83
Partner-rated mate value	100	1.88	5.00	3.60	0.67
Bouts of flu in the past 3 months	97	0	2	0.18	0.41

Bouts of gastroenteritis in the past 3 months	98	0	5	0.72	1.09
Instances of antibiotic use in the past 3 months	98	0	2	0.12	0.36
Illness composite in the past 3 months	97	0	9	2.42	1.78

Women's variables	*N*	Min	Max	Mean	*SD*
Facial femininity	98	1.63	5.54	3.82	0.79
Facial attractiveness	98	1.63	5.15	3.15	0.74
Facial health	98	2.32	5.60	4.35	0.71
Facial dominance	98	2.02	5.90	3.75	0.74
Facial prestige	98	2.29	4.71	3.54	0.60
Handgrip strength	100	19.00	41.50	28.84	4.35
Height	100	150.80	182.30	164.88	6.66
Weight	94	40.00	119.00	57.45	9.70
BMI	94	15.84	45.23	21.20	3.52
WHR	100	0.65	0.91	0.73	0.04
Self-rated mate value	99	1.50	4.88	3.38	0.79
Partner-rated mate value	97	2.00	5.00	3.83	0.67
Bouts of colds in the past 3 months	100	0	5	1.60	1.12
Bouts of flu in the past 3 months	99	0	5	0.30	0.69
Bouts of gastroenteritis in the past 3 months	100	0	5	0.82	1.23
Instances of antibiotic use in the past 3 months	100	0	4	0.32	0.68
Illness composite in the past 3 months	99	0	13	3.04	2.13

Note. BMI = body mass index; WHR = waist-to-hip ratio.

Average ratings for each face on each trait were calculated in the same way as for the British parents sample. For men, facial hair was significantly correlated with all facial traits and for women, makeup use was significantly correlated with all facial traits except facial dominance (*r* > .30 for all significant traits). Therefore, we controlled for facial hair and for makeup use for all traits except women's facial dominance by running linear regressions with facial hair/makeup use as the predictor and each facial trait as the outcome. The saved residuals were used in all analyses of facial traits.

Firstly, in order to identify within-sex trait associations, we ran Pearson's zero-order correlations between all male and all female traits. We then ran separate principal components analyses to identify how traits clustered together within each sex. The saved factor scores were used for hypothesis testing. Using zero-order correlations, we tested the predictions that *i.* there is assortative mating for male and female dimorphism/attractiveness, *ii.* more masculine men have better quality partners, indexed by increased femininity/attractiveness, higher mate value, and better health, and *iii.* more masculine/attractive men and more feminine/attractive women are of better quality themselves, indexed by higher mate value and better health.

5.2.2. Results

Within-sex trait associations. Within-sex zero-order correlations (including information about surviving associations) between facial and anthropometric traits can be viewed in Table 5.3. All correlations will not be summarised in text. However, it can be noted that for men, facial masculinity showed significant positive correlations with facial dominance ($r = .738$, $p < .001$, $q < .001$), handgrip strength ($r = .356$, $p < .001$, $q < .001$), and shoulder circumference ($r = .205$, $p = .043$, $q = .139$). Facial masculinity did not significantly correlate with facial attractiveness (as might be expected if facial masculinity signalled attractive good genes). Facial attractiveness, in turn, showed significant positive correlations with facial prestige and facial health ($r = .744$, $p < .001$, $q < .001$, and $r = .634$, $p < .001$, $q < .001$, respectively), and significant negative correlations with BMI ($r = -.230$, $p = .027$, $q = .109$) and shoulder circumference ($r = -.227$, $p = .024$, $q = .108$). In addition to looking more facially masculine, men with greater handgrip strength also had higher BMI ($r = .409$, $p < .001$, $q < .001$) and more dominant-looking faces ($r = .326$, $p = .001$, $q = .008$), were taller ($r = .311$, $p = .002$, $q = .014$), and looked less healthy ($r = -.214$, $p = .033$, $q = .125$).

For women, all facial traits were strongly and positively correlated with each other (all $r > .50$) except facial dominance. Unlike in men, women's handgrip strength only correlated significantly with height ($r = .382$, $p < .001$, $q < .001$).

Table 5.3

Within-sex Pearson's bivariate correlations between facial and anthropometric traits in the student couples, controlling (where relevant) for men's facial hair and women's makeup use (Study 7). Correlations in women are shown above the diagonal and correlations in men below the diagonal.

	Facial masc/ fem.	Facial attr.	Facial health	Facial dom.	Facial prestige	HGS	Height	BMI	WHR
Facial masc/fem.		**.799***** ***N* = 98**	**.557***** ***N* = 98**	**-.348***** ***N* = 98**	**.631***** ***N* = 98**	-.045 *N* = 98	-.005 *N* = 98	-.228* *N* = 92	-.181 *N* = 98
Facial attr.	.061 *N* = 99		**.681***** ***N* = 98**	-.152 *N* = 98	**.708***** ***N* = 98**	.101 *N* = 98	.012 *N* = 98	-.271** *N* = 92	-.184 *N* = 98
Facial health	.054 *N* = 99	**.634***** ***N* = 99**		-.221* *N* = 98	**.625***** ***N* = 98**	.078 *N* = 98	-.034 *N* = 98	-.250* *N* = 92	-.213* *N* = 98
Facial dom.	**.738***** ***N* = 99**	.090 *N* = 99	.167 *N* = 99		.084 *N* = 98	.069 *N* = 98	.224* *N* = 98	.270** *N* = 92	**.288**** ***N* = 98**
Facial prestige	.191 *N* = 99	**.744***** ***N* = 99**	**.656***** ***N* = 99**	.137 *N* = 99		.114 *N* = 98	.196 *N* = 98	-.219* *N* = 92	-.151 *N* = 98
HGS	**.356***** ***N* = 99**	-.103 *N* = 99	-.214* *N* = 99	**.326**** ***N* = 99**	-.139 *N* = 99		**.382***** ***N* = 100**	-.055 *N* = 94	-.076 *N* = 100
Height	.065 *N* = 99	-.098 *N* = 99	-.079 *N* = 99	.134 *N* = 99	-.113 *N* = 99	**.311**** ***N* = 100**		-.141 *N* = 94	-.141 *N* = 100
BMI	.193 *N* = 92	-.230* *N* = 92	**-.312**** ***N* = 92**	.187 *N* = 92	-.233* *N* = 92	**.409***** ***N* = 93**	-.010 *N* = 93		**.484***** ***N* = 94**
Shoulder circ.	.205* *N* = 98	-.227* *N* = 98	**-.307**** ***N* = 98**	.236* *N* = 98	-.198 *N* = 98	**.515***** ***N* = 99**	**.325**** ***N* = 99**	**.817***** ***N* = 92**	

Note. Attr = attractiveness; BMI = body mass index; circ = circumference; dom = dominance; fem = femininity; HGS = handgrip strength; masc = masculinity; WHR = waist-to-hip ratio.

* $p < .05$, ** $p < .01$, *** $p < .001$. Associations that remained significant after computation of q-values are bolded.

To identify how traits loaded together within each sex, principal components analyses with Varimax rotation were performed separately for men's and women's traits. Factors were saved as new variables. For men, both eigenvalues and visual inspection of a scree plot suggested a four-factor structure (see Table 5.4 for factor loadings). Facial attractiveness, facial health, and facial prestige loaded on the first factor ('facial appeal'); BMI, shoulder circumference, and handgrip strength loaded on the second ('body masculinity'); facial masculinity, dominance, and (to a lesser extent) handgrip strength on the third ('facial dominance'), and height and handgrip strength on the fourth ('stature').

Table 5.4

Principal component loadings for men's traits in the student couples (Study 7).

	Factor			
Trait	Facial appeal	Body masculinity	Facial dominance	Stature
Facial masculinity			**.921**	
Facial attractiveness	**.907**			
Facial health	**.833**			
Facial dominance			**.901**	
Facial prestige	**.891**			
Handgrip strength		**.563**	.316	.415
Height				**.961**
BMI		**.929**		
Shoulder circumference		**.913**		

Note. BMI = body mass index. Factor loadings above .50 are bolded; loadings below .30 are not shown.

For women, a three-factor solution was found (Table 5.5). Similarly to men, all facial traits except facial dominance loaded on the first factor ('facial appeal'). The second factor ('body size') comprised WHR, BMI, and facial dominance, whereas the third factor ('stature') consisted of height, handgrip strength, and facial dominance.

Table 5.5

Principal component loadings for women's traits in the student couples (Study 7).

Trait	Factor Facial appeal	 Body size	 Stature
Facial femininity	**.836**		
Facial attractiveness	**.912**		
Facial health	**.787**		
Facial dominance		**.683**	.416
Facial prestige	**.862**		
Handgrip strength			**.725**
Height			**.836**
BMI		**.740**	
WHR		**.807**	

Note. BMI = body mass index; WHR = waist-to-hip ratio. Factor loadings above .50 are bolded; loadings below .30 are not shown.

Associations between trait factors and indices of partner quality. All between-sex zero-order trait correlations can be seen in Appendix Q. As Table 5.6 shows, we only found one significant correlation between men's and women's trait factors: men with higher body masculinity had partners with greater facial appeal (r = .252, p = .021, q = .102); this association did not survive corrections, however. There was a marginally significant positive correlation between male and female stature (r = .210, p = .055, q = .165), but no other evidence of associations between male and female trait factors. Furthermore, no male trait factors showed significant correlations with either self- or partner-rated female mate value, or with female health. Women who rated themselves as having higher mate value also rated their partners' mate value as higher (r = .293, p = .003, q = .020; this correlation thus did survive corrections). Men and women who rated their partners' mate value as high were also rated by their partners as having higher mate value in turn (r = .228, p = .025, q = .108; non-surviving correlation).

Table 5.6

Pearson's bivariate correlations between trait factors and indices of partner quality in the student couples (Study 7). Male traits are shown in the rows and female traits in the columns.

	Female facial appeal	Female body size	Female stature	Female self-rated MV	Female partner-rated MV	Female colds	Female gastro-enteritis	Female illness comp.
Male facial appeal	.036 N = 84	-.118 N = 84	-.178 N = 84	.096 N = 90	.125 N = 89	.139 N = 91	-.136 N = 91	-.032 N = 90
Male body masc.	.252* N = 84	.115 N = 84	-.017 N = 84	.112 N = 90	.101 N = 89	-.008 N = 91	.091 N = 91	.041 N = 90
Male facial dom.	.165 N = 84	.131 N = 84	.142 N = 84	.171 N = 90	.007 N = 89	.165 N = 91	.001 N = 91	.117 N = 90
Male stature	.120 N = 84	.038 N = 84	.210 N = 84	-.073 N = 90	.083 N = 89	.064 N = 91	.130 N = 91	.134 N = 90
Male self-rated MV	.082 N = 89	-.155 N = 89	.040 N = 89	.161 N = 96	.138 N = 96	.054 N = 97	-.131 N = 97	-.054 N = 96
Male partner-rated MV	.189 N = 92	-.105 N = 92	.066 N = 92	**.293**** **N = 100**	.228* N = 97	.106 N = 100	-.161 N = 100	-.018 N = 99
Male colds	.057 N = 90	-.039 N = 90	.007 N = 90	-.002 N = 97	.105 N = 97	.262** N = 98	.002 N = 98	.265** N = 97
Male gastro-enteritis	-.064 N = 90	-.073 N = 90	-.121 N = 90	.113 N = 97	.126 N = 97	-.065 N = 98	.120 N = 98	.082 N = 97
Male illness comp.	.009 N = 89	-.095 N = 89	-.042 N = 89	-.005 N = 96	.144 N = 96	.180 N = 97	.050 N = 97	.211* N = 96

Note. Comp = composite; masc = masculinity; MV = mate value.
* $p < .05$, ** $p < .01$, *** $p < .001$. Associations that remained significant after computation of q-values are bolded.

Associations between trait factors and indices of own quality. Men with more appealing faces were rated by their partners as having higher mate value ($r = .234$, $p = .026$, $q = .109$) and they also reported having experienced more colds in the past three months ($r = .217$, $p = .040$, $q = .133$). Conversely, taller/stronger men reported fewer bouts of gastroenteritis ($r = -.235$, p =

.026, q = .109). Neither of these associations remained after computation of q-values, and there were no other significant correlations between men's trait factors and indices of quality. Judging by the direction of non-significant effects, however, all attractive/masculine factors appeared to be associated with higher self- and partner-rated mate value. Conversely, there was no consistent direction to the associations between men's trait factors and incidence of common illnesses.

For women, there were no significant associations between any trait factors and quality indices.

Table 5.7

Pearson's bivariate correlations between trait factors and indices of own quality in the student couples (Study 7).

Men's trait factors	Self-rated MV	Partner-rated MV	Colds	Gastro-enteritis	Illness comp.
Facial appeal	.188 *N* = 89	.234* *N* = 91	.217* *N* = 90	-.068 *N* = 90	.102 *N* = 89
Body masculinity	.077 *N* = 89	.195 *N* = 91	-.094 *N* = 90	-.028 *N* = 90	.056 *N* = 89
Facial dominance	.197 *N* = 89	.181 *N* = 91	.038 *N* = 90	.029 *N* = 90	.043 *N* = 89
Stature	.128 *N* = 89	.121 *N* = 91	.081 *N* = 90	-.235* *N* = 90	-.120 *N* = 89
Women's trait factors	**Self-rated MV**	**Partner-rated MV**	**Colds**	**Gastro-enteritis**	**Illness comp.**
Facial appeal	.154 *N* = 92	.181 *N* = 89	.073 *N* = 92	.017 *N* = 92	.062 *N* = 92
Body size	-.153 *N* = 92	.077 *N* = 89	.088 *N* = 92	-.060 *N* = 92	-.008 *N* = 92
Stature	.157 *N* = 92	.001 *N* = 89	.059 *N* = 92	-.076 *N* = 92	.011 *N* = 92

Note. Comp = composite; MV = mate value.
* p < .05. No associations remained significant after computation of q-values.

5.2.3. Interim discussion

Here, we tested the hypotheses that *i.* men and women mate assortatively on dimorphism/attractiveness, *ii.* more masculine/attractive men have better quality partners, indexed by higher mate value and lower incidence of common illnesses, and *iii*. more dimorphic/attractive men and women are of higher quality themselves, using the same quality indices. Overall, we found little evidence for these predictions. Men with more masculine bodies had significantly more facially appealing female partners, as judged by third-party ratings. Interestingly, however, these men's partners did not have higher self- or partner-rated mate value. We found no evidence for the common claim that couples mate assortatively for facial appeal. There was, however, marginally significant evidence for assortative mating for stature. All male masculine/attractive traits were positively correlated with both self- and partner-rated mate value, but only facial appeal showed a significant association. Female traits did not show any significant correlations with mate value (even if female facial appeal was positively associated with both self- and partner-rated mate value, neither association was significant). The only correlation to survive corrections for multiple comparisons was between how women rated their own and their partner's mate value. Lastly, neither male nor female trait factors showed any consistent relationships with common illness prevalence.

5.3. Study 8: British parent couples

5.3.1. Methods

5.3.1.1. Participants

Participants were the British parents from Chapter 4.

5.3.1.2. Materials and procedure

Materials and procedure were the same as described in previous chapters.

5.3.1.3. Statistical analysis

Means and standard deviations for the study variables can be viewed in Table 4.4 (Chapter 4). The analysis followed the same structure as for the student couples: Pearson's zero-order correlations were run to identify both within- and between-sex correlations. Principal components analyses were run to identify within-sex trait clusters. The factor scores were used for correlations testing the hypotheses that *i.* there is assortative mating for male and female dimorphism and attractiveness and *ii*. more masculine men have better quality partners, indexed by increased femininity.

5.3.2. Results

Within-sex trait associations. In men, facial masculinity was strongly correlated with facial dominance (r = .651, p < .001, q < .001), and facial attractiveness was strongly correlated with perceived facial health (r = .869, p < .001, q < .001). No other significant intercorrelations between men's traits were observed (but there was a marginally significant correlation between handgrip strength and height: r = .288, p = .068, q = .191). In women, all facial traits were strongly correlated: facial femininity, facial attractiveness, and perceived facial health were all positively intercorrelated, and facial dominance correlated negatively with all other facial traits (femininity and attractiveness: r = .881, p < .001, q < .001; femininity and perceived health: r = .791, p < .001, q < .001, femininity and dominance: r = -.556, p < .001, q < .001; attractiveness and perceived health: r = .870, p < .001, q < .001; attractiveness and dominance: r = -.434, p < .001, q < .001; perceived health and dominance: r = -.419, p < .001, q < .001). Lastly, taller women had greater handgrip strength (r = .326, p = .002, q = .014). See Table 5.8 for all within-sex trait intercorrelations (including information about which correlations remained after corrections).

Table 5.8

Within-sex Pearson's bivariate correlations between traits in the British parent couples, controlling (where relevant) for age, male facial hair and female makeup use (Study 8). Correlations in women are shown above the diagonal and correlations in men below the diagonal.

	Facial masc/ fem.	Facial attr.	Facial health	Facial dom.	HGS	Height
Facial masc/fem.		**.881***** ***N* = 107**	**.791***** ***N* = 107**	**-.556***** ***N* = 107**	-.020 *N* = 85	-.029 *N* = 84
Facial attr.	.128 *N* = 68		**.870***** ***N* = 107**	**-.434***** ***N* = 107**	.059 *N* = 85	.032 *N* = 84
Facial health	.100 *N* = 68	**.869***** ***N* = 68**		**-.419***** ***N* = 107**	-.017 *N* = 85	-.008 *N* = 84
Facial dom.	**.651***** ***N* = 68**	.034 *N* = 68	.054 *N* = 68		.166 *N* = 85	-.080 *N* = 84
HGS	-.216 *N* = 41	-.009 *N* = 41	.234 *N* = 41	-.034 *N* = 41		**.326**** ***N* = 85**
Height	.009 *N* = 41	.141 *N* = 41	.139 *N* = 41	.007 *N* = 41	.288 *N* = 41	

Note. Attr = attractiveness; dom = dominance; fem = femininity; HGS = handgrip strength; masc = masculinity.
** $p < .01$, *** $p < .001$. Associations that remained significant after computation of *q*-values are bolded.

Separate principal components analyses with Varimax rotation suggested a three-factor solution for men's traits and a two-factor solution for women's traits. Factors, with loadings, can be viewed in Tables 5.9 and 5.10. Men's traits in this more mature sample showed a very similar factor structure as in the younger student sample (the latter including a greater number of traits). Facial attractiveness and perceived facial health loaded on the same 'facial appeal' factor. The other two facial traits, facial masculinity and facial dominance, loaded on the 'facial dominance' factor, which was distinct from the 'stature' structure which comprised height and handgrip strength.

Table 5.9

Principal component loadings for men's traits in the British parent couples (Study 8).

Trait	Factor		
	Facial appeal	Facial dominance	Stature
Facial masculinity		**.895**	
Facial attractiveness	**.953**		
Facial health	**.948**		
Facial dominance		**.880**	
Handgrip strength			**.801**
Height			**.792**

Note. Factor loadings above .50 are bolded; loadings below .30 are not shown.

In women, the two-factor structure was also similar to the three-factor structure solution in the student sample and consisted of 'facial appeal' (showing positive loadings for facial femininity, attractiveness, and perceived health, and a negative loading by facial dominance) and 'stature' (comprising, just like in men, height and handgrip strength).

Table 5.10

Principal component loadings for women's traits in the British parent couples (Study 8).

Trait	Factor	
	Facial appeal	Stature
Facial femininity	**.952**	
Facial attractiveness	**.941**	
Facial health	**.917**	
Facial dominance	**-.656**	
Handgrip strength		**.833**
Height		**.790**

Note. Factor loadings above .50 are bolded; loadings below .30 are not shown.

Associations between trait factors and indices of partner quality. Between-sex zero-order trait correlations can be seen in Appendix R. We identified two significant associations for between-partner trait factors. Unlike in the student sample, couples in this sample mated assortatively for facial appeal (r = .595, p = .001, q = .008; this correlation survived corrections). More facially dominant men also had female partners significantly lower in stature (r = -.377, p = .044, q = .139; did not survive).

Table 5.11

Pearson's bivariate correlations between trait factors in the British parent couples (Study 8). Male traits are shown in the rows and female traits in the columns.

	Female facial appeal	Female stature
Male facial appeal	**.595**** ***N* = 29**	.241 *N* = 29
Male facial dominance	.011 *N* = 29	-.377* *N* = 29
Male stature	.185 *N* = 29	.123 *N* = 29

*** p < .001, * p < .05. Associations that remained significant after computation of q-values are bolded.

5.3.3. Interim discussion

Here, we tested two of the same hypotheses as in the student sample; namely that *i*. men and women mate assortatively on dimorphism/attractiveness, and *ii*. that more masculine men have better quality partners (indexed by increased femininity/attractiveness). Compared to the student sample, this sample was more mature and we had fewer measures overall. We found some evidence of assortative mating here, but not for the prediction that more masculine men have better quality partners. Factor structures in this sample showed similar loadings as in the younger student sample. Between-sex trait factor correlations different from in the students, however. Here, we found evidence for assortative mating for facial appeal, which was absent in the student sample. Furthermore, we also found that more facially dominant men had partners lower in stature, although this was a result unrelated to our predictions and achieved with only this modest sample and so we will not discuss it further at this time. Since men's factor solution here (with fewer measures) did not include body masculinity, it was not possible to assess the significant association that we identified between men's body masculinity and women's facial appeal in the student couples.

5.4. Discussion

In this chapter, across one sample of student couples and one sample of parent couples, we tested the predictions that *i.* men and women mate assortatively on sexual dimorphism/attractiveness, *ii.* more masculine men have better quality partners (indexed by increased femininity/attractiveness, higher mate value, and lower prevalence of common illnesses), and *iii.* more dimorphic/attractive men and women or of higher quality themselves (using the same mate quality measures). Again, we found little support for our predictions.

In the student sample, more bodily masculine men had partners with a more appealing craniofacial structure, although it should be noted that this relationship did not survive corrections for multiple comparisons. We were unable to test this association in the parent couples. If this association would remain in a larger sample, this would support the argument that men with more masculine (i.e., more formidable) bodies are more attractive on the mating market, thus being able to exercise greater choice when selecting a partner. Interestingly, however, the partners of these men did not have significantly higher mate value. It is important to emphasise here that mate value was self- and partner-rated, whereas facial traits were rated by unfamiliar third-party observers. Self-rated mate value is of course mediated by for example self-esteem, whereas partner-rated mate value may arguably by influenced by own self-esteem as well as by relationship quality. While this means that both self- and partner-rated mate value are strongly subjective measures which may be susceptible to a number of biases, they will also be based on true interactions with other possible mates, which might increase the accuracy of these measures. Third-party observers, in contrast, make their attractiveness judgments based solely on a static image of an individual’s face. This finding also opens up to the possibility that third-party ratings of an individual’s craniofacial structure is a poor reflection of that individual’s ‘true’ mate vale in an industrialised society where people engage in extensive grooming and styling behaviour which could considerably enhance their overall attractiveness. The use of static, standardised facial images is justified when trying to assess an individual’s underlying biological quality as indicated by their facial structure, but future research should assess to what extent styling, makeup, clothing, and personality may ‘override’ cues to underlying biological condition.

In the student sample, we found no evidence that couples mated assortatively for facial appeal. This is in contrast to previous research showing that couples tend to be of similar attractiveness (e.g., Feingold, 1988). However, several of the studies included in the meta-analysis by Feingold used participant-provided photographic stimuli, including external cues such as hair and jewellery. Burriss and colleagues (2011), in contrast, used standardised stimuli − and did not observe a significant effect for assortment on facial attractiveness. Again, this *may* be related to the fact that third-party ratings of standardised, masked facial images do not measure someone’s ‘real-world’ attractiveness (perhaps particularly for women, who tend to engage in more styling behaviours than men: Lee et al., 2014; see also Pereira et al., 2019 for similar findings specifically in women, but not men). In a similar vein, Docherty and colleagues (2020) recently reported differing mate preferences as a function of women’s own- versus third-party rated attractiveness. This discrepancy between third-party- and self-rated attractiveness/mate value is corroborated by our finding that our couples did mate assortatively on partner-rated mate value; that is, men who were rated by their partner as having higher mate value also rated their partner as having higher mate value, and vice versa.

Many factors may of course influence this association, but this is an interesting avenue for future research. In the parent couples, however, we did find evidence of assortative mating for facial appeal. Regarding these discrepant findings between our samples, four important points are raised: 1. The student sample was a socioeconomically homogeneous sample, whereas the parents sample was more socioeconomically diverse. It is possible that mating patterns observed in especially ‘WEIRD’ (Henrich et al., 2010) student samples differ from those found in more representative community samples. 2. Simultaneously, however, the student sample size was bigger. 3. The parent sample varied more in age than the student sample. For reasons already explained, we were not able to control for parental age at the time of being photographed. It is possible that the positive association observed between facially appealing traits in the parent couples reflect assortative mating for age (which in turn mediates the appeal of facial traits) rather than actual facial structure. 4. The two individuals in a couple are sometimes claimed to become more facially similar over time (Zajonc et al., 1987), possibly due to similarity of lifestyle/environmental factors (Homish & Leonard, 2008; Silventoinen et al., 2003) and mimicry of facial expressions of emotion, in turn impacting e.g. facial aging in similar ways (Zajonc et al., 1987). This makes it possible that the more mature sample who had been together for a longer period of time (although we did not measure relationship duration in this sample, these couples all had children together and had been together for considerably longer than any of the student couples) showed convergence of facial traits rather than initial assortment on facial appeal (but it should be noted that the phenomenon of convergence of facial traits has been refuted by recent evidence: Tea-Makorn & Kosinski, 2020).

All male masculine/attractiveness trait factors showed positive associations with mate value in the student sample, but only facial appeal showed a significant effect (for partner-rated mate value only). While this latter relationship also did not survive corrections and thus should be interpreted with caution, the overall direction of these associations may suggest that increased masculinity and facial appeal are generally considered to increase men's appeal on the mating market (with the same caveats for mate value as mentioned previously), but future research will have to explore this further. We note, however, that the zero-order correlations between men's traits suggested that more masculine men do *not* look more facially attractive, indicating that a relationship between men's masculinity and their mate appeal is not mediated by increased facial attractiveness per se.

Interestingly, female students' trait factors were not significantly associated with either self- or partner-rated mate value. Zero-order correlations between female traits suggested that more facially attractive women also had a slimmer body (indexed by a lower BMI) but no significant correlation with WHR, which is commonly argued to signal female quality (although we note that the correlation between appealing facial traits and WHR were consistently in the negative direction). Overall, this indicates that women's body size/shape plays less of a role for their value on the mating market than extant literature appears to suggest.

Lastly, in the student sample, we observed no consistent associations between men's or women's traits and prevalence of common illnesses. Facially appealing men had experienced more colds and taller men fewer bouts of gastroenteritis, but overall, the correlations were not consistent in their direction. This is in contrast to common claims in the literature (e.g., the same measures used by Thornhill & Gangestad, 2006; but see also Boothroyd et al., 2013, who reported ambiguous results using these measures), although we should note that we measured a very limited set of self-reported illnesses here. Women's traits showed no significant associations with illness. This suggests that there is no unitary health benefit to increased dimorphism/attractiveness in men or women.

Chapter 6.

General discussion

6.1. Summary of thesis

The overarching goal of this thesis was to investigate sexual selection in humans, with emphasis on selection pressures underlying the evolution of masculine traits in men. In particular, it focused on testing the hypotheses drawn from Evolutionary Psychology that more masculine men should show greater indices of offspring numbers and quality, and greater partner numbers and quality. Given its prevalence in the literature, we focussed particularly on testing predictions derived from the Immunocompetence handicap hypothesis (Folstad & Karter, 1992); namely that masculinity in men is under selection since it signals immunocompetence (although other selection pressures – such as male-male competition – were also considered). Understanding the relationship between sexually dimorphic traits and partner and offspring outcomes is critical in understanding how masculine traits have evolved and/or are being maintained in human males. We also considered other aspects of sexual selection related to sexually dimorphic traits in women as well as attractiveness in both sexes.

6.1.1. Chapter 2

In attempting to understand how/why sexual selection may have shaped the evolution of morphological traits in any species/sex, the primary task must be to demonstrate that such traits are indeed positively associated with biological fitness. Therefore, in Chapter 2, we performed a comprehensive meta-analysis of the associations between masculinity in men’s faces, bodies, voice pitch, height, 2D:4D, and testosterone levels with both reproductive success and mating success. From the extant literature, we included both published and unpublished research from 99 samples, spanning both industrialised ‘WEIRD’ populations and small-scale, traditional societies. The results showed that men’s body masculinity (i.e., greater physical strength, increased muscle mass, and a more masculine body shape) positively predicted both their mating and reproductive success. Effect sizes were weak but robust across both outcome types, different types of study populations, and in published and unpublished works. No other forms of masculinity significantly predicted reproductive success although voice pitch, testosterone levels, and height were associated with significantly greater mating success. Facial masculinity did not significantly predict either mating or reproduction. The lack of significant associations for facial masculinity is especially noteworthy considering that sexual selection research in humans has focused extensively on facial masculinity, particularly on its putative role in signalling men’s heritable immunocompetence, which should thus be selected for under female choice. When interpreting the results of our meta-analysis, it should, however, be noted that the number of observations (and sample sizes within individual studies) varied greatly between traits and outcomes. It therefore cannot be ruled out that we lacked sufficient power to detect significant associations for some traits. Overall, the findings from Chapter 2 support the notion that increased masculinity in men’s bodies – namely greater strength and muscle mass − may be under selection in contemporary human populations. Simultaneously, however, the findings raise questions over the extent to which other masculine traits are subject to current selection.

6.1.2. Chapter 3

Aside from sexually dimorphic traits increasing fitness outcomes due to greater offspring numbers and/or better offspring quality, an additional − but not mutually exclusive − route towards greater fitness is if certain traits increase the chances of acquiring a better-quality mate. Such a mate might then mediate the relationship between the focal trait and reproduction. Therefore, Chapter 3 looked at the relationship between female morphological and attractive traits and reproductive outcomes. This chapter included two new empirical analyses on existing datasets (one of which was from the Agta, a traditional hunter-gatherer society, and the other an urban Turkish sample). The chapter also comprised a systematic literature review considering the links between female traits and reproductive outcomes, and a meta-analysis (including data from our two datasets) of how facial attractiveness in both men and women predicted reproductive outcomes. It was found that amongst the Agta, more facially healthy-looking women had significantly fewer births, while more facially attractive women had better surviving offspring (albeit neither of these results survived corrections for multiple comparisons). In the Turkish sample we detected no significant linear associations. Thus, contrary to expectations, the data both from the Agta women and the Turkish parents implied *lower* offspring numbers for individuals with more appealing facial traits (which is the opposite to what we would expect, if such facial traits communicated greater fertility). The systematic review of maternal traits predicting reproduction comprised measures of 2D:4D, voice pitch, and upper-body strength in women from 24 samples, and showed a mixture of null, positive, and negative effects. Given how commonly traits such as WHR and breast size are claimed to signal fertility in the literature, it is noteworthy that we were not able to locate a single study measuring fitness as a function of either of those traits. The meta-analysis portion, including observations from seven samples in each sex, showed no significant associations between facial attractiveness and fitness outcomes. For samples of non-contracepting women, the meta-analytic effect of r = .156 (N = 319) suggests a significant effect may be possible with a larger N/k, but the dearth of available studies means it is very hard to assess such a relationship. For men, effect sizes centred around zero in both contracepting and non-contracepting samples (N = 770 and 216, respectively). While nonsignificant effects should be interpreted with caution, these findings do not lend support to the notion that male facial attractiveness is under selection in humans.

6.1.3. Chapter 4

In Chapter 4, we then looked at another key prediction of the ICHH: that if sexual dimorphism signals heritable quality, children of more sexually dimorphic parents should be of better quality themselves. We tested this prediction in a U.K. sample of families as well as the Turkish families from Chapter 3. Quality was indexed by better health (i.e., reported prevalence of common illnesses, sampled in the British children at three timepoints – once in late childhood and twice in adolescence − and during the first five years of life in the Turkish children) and earlier romantic interest/sexual onset (where earlier onset may lead to greater reproductive success; this was self-reported by the British adolescents). Very few significant associations between parental traits and offspring health were observed. The majority of these were for paternal traits and in the predicted direction (all of which were for facial dominance, handgrip strength, and height, and none for facial masculinity/attractiveness). Teenagers of more facially attractive and healthy-looking fathers showed earlier sexual onset. In contrast, stronger mothers had offspring reporting lower interest in romantic relationships, as well as one measure of worse health. None of these survived corrections for multiple comparisons, however. For the Turkish families, there were no significant associations even before corrections were applied. Meta-analysing the effects on offspring health from the two samples in this chapter together produced overall effect sizes close to zero for both paternal and maternal facial traits. These findings do not support the notion that parental traits are reliably associated with offspring's experiences of common illnesses, nor with sexual onset in offspring.

6.1.4. Chapter 5

In the final empirical chapter, we followed up on the question raised in Chapter 3 − whether masculine men acquire better quality partners − by testing partner quality associations in a sample of U.K. student couples as well as the British sample of parents from Chapter 4. Male students who had more masculine bodies (based on a principal component of BMI, handgrip strength, and shoulder circumference) had more facially appealing (attractive, feminine, prestigious- and healthy-looking) partners. Male students with greater facial appeal (more attractive, prestigious- and healthy-looking) were rated by their partners as having higher mate value, but they also self-reported more − rather than the expected fewer − colds. In contrast, male students with increased stature reported fewer bouts of gastroenteritis. Student couples also showed assortative mating for partner-rated mate value, and females who considered their own mate value to be high also rated their male partners as having higher mate value. The only correlation of these to survive corrections for multiple comparisons was the last one. Notably, we found no evidence here of assortative mating for facial appeal. Conversely, amongst the parent couples, we did observe a positive association between how facially appealing (for men: attractive and healthy-looking; for women: attractive, feminine, healthy-looking and not facially dominant) the partners were in each couple. This was the second result in the chapter to survive correction for multiple comparisons. More facially masculine/dominant-looking men also had partners significantly lower in stature/strength. Body shape/size was otherwise not measured in the parent couples. Taken together, these findings suggest that men with more masculine bodies may have more facially attractive female partners, and that greater facial appeal in men *may* be associated with some indices of greater mate value – but the evidence for the latter is not clear.

6.2. Key themes

6.2.1. Very few significant results

A key theme across this thesis is how very few significant findings we detected. This pattern persisted across all research questions tested across all four empirical chapters. Given the frequency and the fervour with which sexually dimorphic traits in both men and women are argued in the sexual selection literature to communicate quality indices (e.g., Johnston & Franklin, 1993; Gülcen et al., 2021; Rhodes et al., 2003; Thornhill & Gangestad, 2006), it could be expected that these claims would be well-established and supported by empirical evidence. It is therefore quite remarkable just how few significant findings we observed. A caveat is that some of our samples were very small and likely underpowered, and associations would have needed to be very strong in order to be significant. The meta-analysis in Chapter 2 showed that the robust associations between men's body masculinity and fitness indices showed correlation coefficients of approximately .15; an effect size of that magnitude would require a sample size of 172 participants to be significant. Here, most analyses had sample sizes smaller than that, and it cannot be ruled out that some of the nonsignificant findings here would have been significant with greater power. However, this thesis also included two meta-analyses based on systematic literature searches, with sample sizes sometimes reaching the thousands; these analyses still yielded remarkably few significant effects. Notably, some of the meta-analytic associations showed effect sizes close to zero (e.g., men's facial attractiveness predicting reproductive outcomes, with $N > 1{,}900$ participants). Across the entire thesis, the total sample of unique participants (including the meta-analyses) equalled more than 179,000 men and 1,800 women, and the total number of analyses 1,150. Of these, 97 significant associations remained significant after computing q-values. If the associations between sexually dimorphic traits and quality indices were anywhere near as strong as much of the current literature appears to suggest, one might expect that more than eight percent of analyses would have yielded significant associations.

6.2.2. No significant effects for facial masculinity

As mentioned previously in this thesis, men's facial masculinity has been subject to considerable attention in the sexual selection literature, particularly as a quality-signalling trait argued to be selected for under female choice. Here, men's facial masculinity did not significantly predict any aspect of their mating or reproductive success, offspring's prevalence of common illnesses, offspring's age at sexual onset, partner's quality indices (sexually dimorphic/attractive traits, common illnesses, or mate value), or own quality (common illnesses and mate value). It should be noted that, in the meta-analysis in Chapter 2, some of the associations for mating success and reproductive success were marginally significant prior to corrections for multiple comparisons, but with correlation coefficients < .10. Across the entire thesis, facial masculinity showed a single significant effect: as part of the principal component 'facial dominance', facial masculinity was significantly associated with lower stature/strength of female partners in British parent couples. This effect was not predicted, and also did not survive corrections for multiple comparisons. Again, it is remarkable that facial masculinity does so little, in spite of so large a literature claiming the opposite (e.g., Ditzen et al., 2017; Foo et al., 2020; Gangestad & Simpson, 2000; Gangestad & Thornhill, 2003; Penton-Voak & Perrett, 1999; Rhodes, 2006; Rhodes et al., 2003; 2005).

One important point is that if men's facial masculinity does function as a 'handicap' signalling their genetic quality to potential mates, we might not necessarily expect it to be associated with better health per se (Getty, 2002) – the key thing here might rather be that men high in this 'handicap' do not suffer disadvantaged health. An issue with this notion, however, is that traits that function as handicaps confer some kind of cost on the bearer (Zahavi, 1975; Zahavi & Zahavi, 1999). Facial masculinity remains stable post-development and is (unlike other handicap traits such as the colourful ornaments exhibited by males in some bird species: Hamilton & Zuk, 1982) not related to current parasite burden nor is it metabolically costly to maintain. It is therefore not clear how strong the rationale is for arguing that it functions as a condition-dependent ornament.

Why do we see sexual dimorphism in human craniofacial structure, if facial masculinity is not under present selection? There are several possible explanations for this.

One is phylogenetic inertia, whereby a trait was under selection in a species/sex in the past but is no longer under selection in the present (Cheverud et al., 1985). Provided that a previously evolved trait is not too costly to develop and maintain, it may be retained even in the absence of selection pressures acting on it.

Another possible explanation is allometry: the tendency for increased height to result in different body proportions (Gould, 1966). In a species like humans where one sex is taller than the other sex (Gray & Wolfe, 1980), allometry might contribute to the taller sex exhibiting a facial morphology differing from that of the shorter sex (Mitteroecker et al., 2013; Zaidi et al., 2019). Simultaneously, however, evidence suggests that facial dimorphism is not solely due to allometry (Kleisner et al., 2021; Schaefer et al., 2004; Weston et al., 2007); this indicates that testosterone influences the development of facial masculine features independently of men’s increased height. Therefore, human facial dimorphism is not likely to be the result of allometry alone, even if it does play a role.

Another possible explanation, which has received attention in the literature, is that facial masculinity is an indirect cue to a man’s formidability/competitiveness. In a species where males engage in intrasexual competition (which is likely to have been the case in human evolution: Hill et al., 2017; Gat, 2015), it would arguably be beneficial to be able to assess a potential rival’s formidability (Sell et al., 2009). In both our student sample and the British parents sample, facial masculinity was very strongly correlated with rated facial dominance, where dominance was defined as the likelihood that the individual in the image would win a fistfight with an age-matched, same-sex individual. The relationship between facial masculinity and perceived dominance is well-established in the literature (e.g., Boothroyd et al., 2007; Fink et al., 2007). Zero-order correlations in the student sample also showed that facial masculinity was moderately correlated with handgrip strength (although this was not replicated in the parent sample). Our findings in the student sample corroborate previous findings that people treat a man’s craniofacial structure as a cue to his strength and overall formidability. For example, Windhager and colleagues (2011) demonstrated that in a small Western sample, men with greater morphometric facial masculinity had higher handgrip strength and were also rated as looking more dominant and masculine (but not more attractive). Similar findings have been reported elsewhere for both Western samples (rated facial masculinity: Fink et al., 2007; but see Holzleitner & Perrett, 2016, who only detected a weak relationship between men’s actual and perceived strength judging from 3D face shape) and in the Masaii pastoralists (morphometric facial masculinity: Butovskaya et al., 2018). ‘WEIRD’ observers can also accurately assess strength from facial photographs alone, both for own-ethnicity and other-ethnicity faces (Sell et al., 2009). Such judgments are made extremely rapidly (Albert et al., 2021b), suggesting that this process is automatic.

In terms of more objective measures of physical competitiveness, Kordsmeyer and colleagues (2019) found that more dominant-looking German men were more likely to win a physical competition (arm wrestling) but not non-physical competitions; this effect was mediated by greater upper-body (but not handgrip) strength. Little and colleagues (2015) also reported that third-party observers were able to pick the winners from men’s Mixed Martial Arts fighting competitions, judging from facial images of the fighters (although this was not replicated by recent evidence from Třebický et al., 2019). Winners’ faces were also rated as looking stronger, more aggressive, and more masculine. It should be mentioned here that this is an example of a very specific sample, however, possibly located at the extreme end of formidable appearance.

Furthermore, the relationship between facial dimorphism and physical strength might be stronger in men than in women: in both our U.K. samples, women’s facial traits were unrelated to handgrip strength. Our results are similar to observations reported recently by Hahn and colleagues (2019) between women’s morphometric facial masculinity and handgrip strength in a large sample of U.K. women. Sell and colleagues (2009) found that third-party observers were more accurate at assessing male than female strength from facial cues alone. However, in the Masaai pastoralists, female facial shape was significantly correlated with handgrip strength, with correlations similar in magnitude to in men (Butovskaya et al., 2018). As the authors point out, genetics may influence handgrip strength more in men and environmental factors may be more influential in women (Isen et al., 2014), which may explain the contrasting relationships observed in industrialised and non-industrialised populations.

While male-male physical aggression does persist in humans (Archer, 2019; Blaker & Van Vugt, 2014), it is undoubtedly relatively infrequent in contemporary human societies, which are generally very tolerant and cooperative (Boyd & Richerson, 1982; Kaplan et al., 2005). One may therefore question the adaptive value of being sensitive to formidability cues in a species where violence is relatively rare. However, human societies are ultimately strongly hierarchical (Anderson et al., 2015), and facial dominance may positively predict status and resource holding potential in such hierarchies (Parker, 1974). For example, Muller and Mazur (1997) observed that facially dominant-looking men in the U.S. military were more likely to achieve higher career success. Similarly, more dominant-looking political candidates may be preferred by voters, at least in times of conflict (Little et al., 2007; Spisak et al., 2012), and companies run by more facially dominant CEOs make greater financial profits (Rule & Ambady, 2008). In a similar vein, Doll and colleagues (2014) found that third-party ratings of fraternity men’s faces predicted both their peer-rated physical formidability and leadership abilities. Thus, even in comparatively pacifist contemporary human societies, more physically formidable men (more so than formidable women) appear to have an advantage when climbing status hierarchies (Blaker & Van Vugt, 2014). Both Blaker and Van Vugt (2014) and Lukaszewski and colleagues (2016) suggest that humans possess adaptations to help us detect and navigate such hierarchies. For example, Lukaszewski and colleagues argue that humans readily confer status to men with increased physical formidability, since such men are perceived as able to benefit group welfare by effectively negotiating ingroup relations while protecting the ingroup against outgroups. This might explain why men with more formidable-looking faces have an advantage when navigating status hierarchies in socially complex societies. Thus, facial formidability – which is strongly correlated with facial masculinity – may function as a proxy for overall formidability, rather than being *directly* selected for.

6.2.3. Body masculinity in men may be important

The previous point leads to the next key theme: that we do find evidence in this thesis that body masculinity – indexed by physically formidable traits such as increased strength, muscle mass, and a more masculine body shape – is weakly associated with some fitness measures. The significant associations we detect here were for mating and reproductive success; for a *very* limited number of offspring health outcomes (none of which survived corrections); and for facial appeal of these men’s partners (which also did not survive, but which did show a moderately-sized correlation).

Overall, our findings thus suggest that formidable traits in men’s bodies might be under weak selection in contemporary humans. What selection pressures are likely to underlie this? Typically, formidable traits are selected for under male-male competition. Greater physical strength is arguably advantageous in direct contests, and increased body size likewise confers competitive advantages, both through its association with strength and through providing ‘buttressing’ against impact, as well as by intimidation in indirect displays (Plavcan, 2012). Our closest primate relatives show greater degrees of sexual dimorphism (largely consisting of size and canine dimorphism) than humans do, and this is likely to have evolved through intrasexual competition (Plavcan, 2012). Male primates use physical dominance to gain priority access both to food resources and to females (Boesch et al., 2006; Cowlishaw & Dunbar, 1991), and male physical size is strongly linked (*r* ~ .70) to social status in primate societies (von Rueden & Jaeggi, 2016). The stronger such intrasexual competition, the stronger the selection pressures on sexually dimorphic traits: sexual dimorphism is typically (but not always) greatest under polygyny, moderate in multi-male, multi-female groups, and minimal under monogamy (Clutton-Brock & Harvey, 1977). In comparison, human size dimorphism is slightly greater than in strictly monogamous primate species (Dixson, 2009), and in non-industrialised human populations, the correlation between male body size and social status is positive but considerably weaker than in non-human primates (*r* ~ .20: von Rueden & Jaeggi, 2016). The ancestral prevalence and importance of male aggression, violence, and intergroup conflict has been discussed in the literature (e.g., Goetz, 2010; Walker, 2001), but clearly must be reconciled with the fact that humans have also evolved to become a remarkably cooperative and tolerant species (Boyd & Richerson, 1982). Overall, this suggests relaxed but maintained selection pressures for formidability in human males. As outlined in the previous subsection, physical formidability appears to confer men greater dominance, status, and resources (von Rueden et al., 2008). Importantly, this pattern is not limited to small-scale traditional societies but is present in ‘WEIRD’ populations as well (Lukaszewski et al., 2016). Moreover, unlike morphological traits that remain stable post-pubertal development (such as facial masculinity, which shows, at best, a very weak correlation with adult testosterone levels: Kordsmeyer et al., 2019; Peters et al., 2008), strength and muscle mass are continuously dependent on adult testosterone levels (Gettler et al., 2010). Since testosterone also motivates competitive and status-enhancing behaviours (reviewed in Eisenegger et al., 2011; Hamilton et al., 2015; and Mehta & Prasad, 2015), this means that it is also possible that the relationship between male formidability and status is, in part, mediated by adult testosterone levels.

Taken together, extant evidence seems to go hand-in-hand with the evolution of a species subject to self-domestication, with less emphasis on male-male physical competition and more focus on in-group cooperation (in addition to the evolution of pair-bonds, biparental care, and mutual mate choice, relaxing intrasexual mate competition further). This would have led to reduced selection pressures on male formidability (resulting in reduced levels of sexual dimorphism: reviewed in Cieri et al., 2014), but where formidability still functions as a pathway towards increased status. Female partners of formidable men may thus gain both direct benefits in terms of increased access to resources/protection, and indirect benefits in terms of increased chances of surviving offspring and possibly more competitive sons, resulting in increased inclusive fitness. Thus, the relationship between physical formidability and fitness may therefore be mediated by status. If this is the case, women might preferentially choose men based on their competitiveness for status, and/or men may acquire increased access to female partners through their higher status (as suggested by Scott et al., 2014). This is further corroborated by evidence that the relationship between men's formidable traits and mating success is mediated not by increased female-rated attractiveness, but by male-rated dominance (Hill et al., 2014; Kordsmeyer et al., 2018). Future research should explore the exact mechanisms mediating the relationships between male formidability, status, and female mate choice.

Interestingly, increased height is often proposed as an aspect of physical formidability, but here we did not find that height increased men's fitness indices. This is noteworthy, since status benefits of being tall are well-established (reviewed in Blaker & Van Vugt, 2014). However, it has also been proposed that, while the associations people make between body size (i.e., muscularity) and dominance have been suggested to be adaptations, appearing very early in life (Thomsen et al., 2011), the associations between height and status might rather be culturally learnt (Blaker & Van Vugt, 2014). This also appears supported by the fact that in

Western populations, women express clear preferences for, and choose, male partners taller than themselves (the ‘male-taller norm’: Stulp et al., 2013), but this pattern might be culture-specific (Sear & Marlowe, 2009).

6.2.4. What do women’s traits communicate?

Feminine traits in women, such as a feminine craniofacial structure and a more curvaceous figure, are very commonly argued to signal health and fertility (reviewed in e.g., Bovet, 2019). Across this thesis, we found no associations between female traits and any fitness indices that survived corrections. In our own datasets, we detected a limited number of significant associations suggesting that more facially appealing women had fewer – rather than more – offspring. The direction of the non-significant meta-analytic effect rather suggested that overall, there *may* be a positive correlation between female facial attractiveness and offspring numbers in non-contracepting populations only. In Chapter 3, we argued that future research should investigate, firstly, whether this association holds in larger samples, and secondly, whether this is mediated by age at marriage (which would not indicate greater fecundity per se). We found no indication that any of the female traits we investigated predicted offspring health, offspring sexual onset, women’s own health or mate quality, or the health or quality of their partners. The only significant effects we found for women was, as already mentioned, that more facially appealing women had male partners with more masculine bodies, and that women who rated their own mate value as higher rated their partner’s mate value as higher as well.

These findings are challenging to reconcile with how commonly women’s facial femininity/attractiveness are argued to reflect their quality. Why might people be attracted to facial femininity if it does not communicate fitness indices? Given that we are attracted to neotenous features in female faces (Perrett et al., 1994; Rhodes, 2006) which make faces look both feminine and young – and youth *is* important for women’s fertility (e.g., Menken et al.,

1986) – it is possible that our preferences for femininity/neoteny largely correspond to preferences for youth. Additionally, neotenous and/or feminine features elicit nurturant responses (Cunningham, 1986, Cunningham et al., 1990), make faces look more prosocial (Barber, 1995), and are also associated with greater maternal tendencies (Law Smith et al., 2012); all of which are traits that may be favoured in women despite not being related to greater reproductive potential per se.

The most striking finding from the systematic literature search for Chapter 3 was how very little empirical evidence there is assessing the associations between feminine traits and fitness outcomes. For example, we did not locate a single study measuring reproductive outcomes as a function of facial femininity. Given that facial femininity and facial attractiveness are strongly correlated, they may be treated as measuring largely the same thing. However, a caveat in this area is that, as expressed by Scott and colleagues (2014), most research is conducted in 'WEIRD' samples, with evidence from traditional, small-scale societies not necessarily supporting the notion that facial femininity is always or universally attractive to men (e.g., Boothroyd et al., 2021). Overall, the theoretical grounds for facial femininity to be an evolved signal communicating fertility are weak: as mentioned in the Chapter 3 introduction, unlike facial masculinity in men, facial femininity in women does not become exaggerated at sexual maturity and is thus technically not a secondary sexual characteristic. While the fuller lips seen in women are often claimed to be oestrogen-dependent (e.g., Rhodes, 2006), the evidence for a relationship between oestrogen and other feminine facial traits is scarce. Combined with the fact that female faces cease to grow in puberty (Bulygina et al., 2006) – when oestrogen levels are high and stimulate exaggeration of other feminine traits such as enlarged breasts and a gluteofemoral fat distribution – suggests that a feminine craniofacial structure is not dependent on oestrogen. Rather, human facial dimorphism appears to largely result from testosterone exposure influencing the growth of males' faces (Swaddle & Reierson, 2002), further weakening the claim that facial femininity in women is a signal to communicate reproductive potential.

As mentioned above, whereas femininity in women's faces is not sensitive to pubertal oestrogen exposure, female body shape (particularly breast size and WHR) does change in adolescence (although it is unclear to which extent women's body shape corresponds to adult oestrogen levels: Grillot et al., 2014; Jasieńska et al., 2004; Jones et al., 2018; Mondragón-Ceballos et al., 2015; Rilling et al., 2009). It can therefore be argued that there is a stronger theoretical foundation for assessing women's fitness in relation to their bodily rather than facial traits. Unfortunately, however, we were not able to locate any studies which had measured women's reproductive outcomes as a function of pre-reproductive breast size or WHR. Clearly this is an interesting avenue for future research.

Overall, there exists a widespread tendency to focus almost exclusively on female fertility in human sexual selection research; however, successfully raising an offspring to adulthood is no less important to an organisms' biological fitness than producing that offspring in the first place (Burch, 2020) – and mothers and other female kin are especially important to offspring survival (Sear & Coall, 2011). There is therefore no reason to expect that phenotypic traits suggestive of good maternal qualities may not be prioritised at least to a similar extent as cues to fertility are, particularly in a pair-bonding species like humans. We suggest this to be investigated further in future research.

6.3. Current limitations and future directions for human sexual selection research

6.3.1. Methods and measures

Facial dimorphism. In this thesis, we focussed primarily on the traits facial dimorphism, handgrip strength, and body size/shape as predictors of fitness indices in men and women.

Whereas handgrip strength and body measurements are easily obtained and commonly used measures, the same cannot be said for facial dimorphism: this may be measured through subjective third-party ratings; through more objective geometric morphometric dimorphism scores (although there is an element of subjectivity in how ‘landmarks’ are placed on facial images for morphometric analysis); and through facial measurements such as fWHR – where the latter is sometimes done through measurements from facial images and sometimes from actual craniofacial measurements taken directly on dry skulls. Whereas these different ways of quantifying facial dimorphism may correlate, they do not correlate perfectly, and each are subject to weaknesses (discussed in Sanchez-Pages et al., 2014). It is also not clear to what extent these different measurement types are reliably correlated with developmental exposure to sex hormones, or indeed if some of them are even sexually dimorphic in the first place (Lefevre et al., 2012; Mitteroecker et al., 2015). In terms of the general research area, the diversity of measurement methods and the lack of a ‘gold standard’ for assessing facial dimorphism are problematic and should ideally be rectified.

Here, we opted for third-party facial ratings. The rationale for using ratings is that facial dimorphism is not likely to be an evolved signal if people are not capable of detecting it in the first place. However, there are several shortcomings to this method: despite attempts to standardise images across participants, there may nonetheless be subtle differences in, for example, participants’ facial expressions and head tilt, which may influence how their images are rated (Witkower & Tracy, 2019). Furthermore, since the data collections we did here were part of larger data collections where we had to prioritise sample size, we were not able to require participants to completely remove makeup and jewellery, which would have been desired. The student couple data collection also took place largely in November, when several male participants had grown partial facial hair in support of the annual male health awareness campaign ‘Movember’ (an admirable endeavour, but not ideal from a facial stimuli quality perspective). The facial stimuli from the archival datasets (the Agta and the Turkish samples) were not fully standardised, but we nonetheless considered these datasets valuable additions. For the standardised U.K. facial stimuli, our intention was to also complement facial ratings with geometric morphometric scores; however, the analysis failed to distinguish between male and female faces and we therefore chose only to include facial ratings.

Fitness outcomes and indices. As raised in Chapter 2, when testing human fitness outcomes, researchers use a wide range of measures spanning from preferences for casual sex or intended infidelity, to actual fitness outcomes such as number of offspring or offspring survival. Such diversity of outcomes is, to an extent, inevitable when testing samples that vary enormously in, for example, mating patterns (polygyny versus serial monogamy versus strict monogamy), contraceptive practices, and access to medical care. However, as we have shown, some widely used proxies for fitness may not test what they are intended to test: for example, in Chapter 2, we found that several masculine traits significantly predicted mating outcomes (primarily in 'WEIRD' samples) but did not predict reproductive outcomes (largely tested in traditional samples). If proxies for fitness (i.e., mating) were robust, we would expect to see similar patterns across mating and reproductive outcomes. The fact that we do not always do that may be because these proxies are poor proxies to begin with – but may also stem from the considerable diversity we see in mating patterns between different populations. In large-scale industrialised populations, where people are exposed to a seemingly endless number of potential mates, certain individuals may be able to act out, for example, preferences for casual mating in a way that would have been impossible in a small-scale or ancestral environment. Studying human mating behaviour in industrialised populations may thus allow us to more clearly see underlying adaptations playing out when relatively free of ecological constraints. However, when we oversample from 'WEIRD' samples for sexual selection research, this may simultaneously skew our perception of how strongly certain traits are preferred and/or selected for.

Temporal aspect. Another issue in sexual selection research relates to the temporal aspect of when both predictor traits and outcomes are measured. Sexually selected traits develop primarily under the influence of sex hormones in adolescence. Such traits do not always correlate highly with adult levels of hormones but are sometimes treated as if they do so (e.g., facial masculinity). When traits are treated as proxies for an individual's current hormone levels, but *actually* reflect that individual's hormone levels when they were sexually maturing, that is not an informative predictor. It would undoubtedly be preferred if longitudinal designs were used to a greater extent, as this would ultimately be more informative (e.g., Foo et al., 2020). A related issue, specifically when testing biological fitness, is that reproduction influences the expression of certain traits. This is particularly the case for women, whose hormone-mediated bodily traits (WHR and breast size) are permanently altered by childbirth. It is imperative to measure predictor traits prior to reproduction has commenced; again, this would require longitudinal designs.

6.3.2. Sample sizes and study populations

One of the key themes in this thesis was the lack of significant results we detected. While our sample sizes differed considerably between analyses, it should also be noted that sample sizes were in some of our analyses very small. This is not unusual in the current literature and, given that the significant associations we did find were weak, it is imperative that researchers aim to recruit larger samples. Here, we attempted to partially solve the issue of small sample sizes by meta-analysing effects across samples. The need for greater sample sizes also, however, needs to be reconciled with attempting to combat the overreliance on 'WEIRD' samples that is common in evolutionary research (Pollet & Saxton, 2019) and increase sampling from more diverse samples. Oftentimes, however, this also constrains sample sizes, particularly when recruiting from small-scale populations.

6.4. Implications for human sexual selection models

Human sexual selection models often focus exclusively on one type of selection (e.g., female choice), neglecting to take into account competing or alternative pressures. This appears to have resulted in a fairly singular view on the evolution of human sexual dimorphism, where certain types of selection pressures have received considerable attention in the literature at the cost of others (reviewed in Puts, 2010). The human lineage appears likely to have originated from a more strongly sexually dimorphic, polygynous ancestor where male-male competition played a greater role (although there is a degree of speculation here: Plavcan, 2012), undergoing considerable changes to arrive at current levels of comparatively moderate dimorphism. This reduction in dimorphism was likely mediated by the evolution of greater social tolerance and cooperation (self-domestication) as well as pair-bonds, extensive biparental care, and mutual mate choice and a relatively flexible mating system, switching between occasional polygyny (Flinn & Low, 1986) and more prevalent (typically serial) monogamy (Walker et al., 2011). These patterns make humans a highly unusual primate – and reinforces the argument that it is not useful to consider either female mate choice or male-male competition in isolation when considering the evolution of extant human dimorphism. Current sexual selection models also typically fail to consider the multitude of other factors that influence human mate choices, such as assortment on sociodemographic variables and personality traits (Botwin et al., 1997; Luo & Klohnen, 2005). It is unclear to what extent assortment on such factors can be reconciled with evolutionary models which almost exclusively emphasise mate choice done on fitness-communicating morphological traits. An additional caveat in human sexual selection research is that human mating and marriages are negotiated (or at least influenced) by parents and other kin in many cultures (and this practice likely existed in ancestral human populations as well). This might prevent individuals from freely acting on their mate preferences and constraining mate choice (Walker et al., 2011), although that is not to say that arranged marriage practices may not co-exist with mate choice (Fletcher et al., 2015).

As a final note on the relationship between human facial dimorphism and fitness benefits, there appears to be a tendency in the human sexual selection literature to assume that exaggerated sexual dimorphism in both male and female faces confers benefits. This would imply that both men and women have been under sexual selection pressures to evolve facially dimorphic traits. Typically, sexually dimorphic traits are traits that are exaggerated in one sex as a result of sexual selection pressures specifically on that sex, not the other. An analogy explaining human strength dimorphism would be to argue that while men have faced selection pressures to become strong, women have faced selection pressures to become weak. It is certainly the case that opposing pressures can operate on the two sexes, but these are typically in the form of sexual selection exerting pressures on one sex with natural selection exerting pressures on the other. An example of this would be the males of a species evolving ornaments to attract females while females remain dull in appearance, likely to evade detection by predators (e.g., Jordão & Oliveira, 2001). This is what we would expect if the male trait evolved as a costly handicap – it would be selected *against* in females. It is unclear whether there are reasons to argue or expect men's and women's faces to have faced opposing sexual selection pressures which would then result in facial dimorphism conferring fitness benefits for both sexes, as is claimed in the literature – particularly when the prevalent view in the literature appears to be that male facial masculinity may be a costly handicap (however weak the empirical evidence for that claim actually is).

6.5. Conclusion

The themes emerging from this thesis suggest that traits associated with male formidability, such as increased strength and muscle mass, might remain under sexual selection in humans, as evidenced by robust positive correlations with both mating and reproductive outcomes. It is imperative to note, however, that these selection pressures are likely considerably weaker in humans compared to in our closest primate relatives. Moreover, physical formidability may, at present, be unlikely to be selected for under direct physical contests – as is common in other human mammals – but might rather operate primarily via increased social status/resource holding potential (even if direct contests may also occur). Also, as indicated by how weak the associations between physical formidability and fitness indices are in human males, male-male competition is likely to exist in concert with other selection pressures – but we do not find compelling evidence that female choice for male heritable *immunocompetence* is likely to be one of them. Taken together, two important conclusions follow from this thesis: *i*. Arguments for sexually dimorphic traits communicating heritable immunocompetence and/or fertility are considerably weaker than presumed – and for some traits, may be lacking altogether. This illustrates the need for associations to be tested directly rather than relying on assumptions. *ii*. Going forward, sample sizes need to be considerably bigger and samples more diverse, particularly when testing associations that are likely to be weak.

Overall, the findings emerging from this thesis call for a more nuanced view of sexual selection in humans, taking into account the socially complex and – compared to other non-human mammals – highly flexible mating conditions under which modern humans have evolved and continue to exist.

Appendix A

Effect size conversion formulas

Kendall’s tau	$r = \sin(.5\pi\tau)$ (Kendall, 1970)
Spearman's rho	not converted
t	$rY\lambda = \sqrt{(t^2 / (t^2 + df))}$ Online converter: https://www.uccs.edu/lbecker/
Odds ratio	Online converter: http://escal.site/
Unstandardized regression coefficient (B)	β = (S.D. of predictor/S.D. of outcome) × B; $r = .98\beta + .05\lambda$ ($\lambda = 1$ when β is nonnegative and 0 when β is negative)
Standardized regression coefficient (β)	$r = .98\beta + .05\lambda$ ($\lambda = 1$ when β is nonnegative and 0 when β is negative; Peterson & Brown, 2005)

Appendix B.1

General coding decisions

Sexual orientation	Coded as non-heterosexual sample if the sample was mixed but predominantly heterosexual.
Samples or subsamples comprising only fathers and/or married individuals	Coded as heterosexual unless otherwise specified, since they had reproduced/married heterosexually.
Student samples with a mean age ≤ 20	Coded as non-fathers.
Sample contained ≥ 50% students	Coded as a student sample.
Sample contained both students and non-students but the proportion of students/non-students was not mentioned	Coded as a non-student sample.
Age	Considered an essential control for all outcome variables except age at first sexual intercourse/encounter and age at the birth of the first child, unless all participants were the same age, and for mating attitude measures.
Ethnicity	Coded as 'white' if ≥ 75% of sample was white.
Marriage system	Coded as polygynous if polygyny was permitted in population, even if rare.
Online samples	Coded as low fertility and monogamous.
Cut off point for high versus low fertility	3.0 children/woman (in sample or population at the time of sampling).
Extreme outliers	Were included when possible, as outliers were expected.
Analyses of relevant relationships were included in paper, but authors had submitted results/raw data to us (e.g., results for men only, controlling for age etc.).	Coded as published results.
Preprints	Coded as non-published and non-peer-reviewed, unless the paper was later accepted for publication in which case it was updated as published and peer-reviewed.
Paper contained both zero-order correlations and multiple regression coefficients	We chose the regression coefficient if the multiple regression included relevant control variables (such as age), and the correlation coefficient if the multiple regression included irrelevant control variables.
Effect sizes given as Spearman's rho	Were not converted; however, were coded as converted for moderation analyses, because it was not given as *r* and therefore considered to be an estimate.
Number of children in industrialised populations	Coded as children born (rather than surviving children) unless otherwise specified. (In naturally fertile populations, it is typically

	spelled out whether measures refer to number of children born vs number of surviving children.)
Dataset on age at first sexual intercourse contained virgins	Current age was used.
Testosterone studies where the authors only included samples that were clear	Coded as having controlled for blood contamination.
Muscularity measures	When other-rated, adiposity should be controlled for and was thus considered a necessary control; when own-rated, adiposity was not considered a necessary control, since people should be able to assess their own amount of muscle/adiposity.
Handgrip strength	Moderator ‘number of measurements’ was coded as number of measurements per hand, not in total.

Appendix B.2

Study-specific coding decisions

Authors	Decisions
Alvergne et al., 2009	We assumed N=53 (married fathers only) as *p*-value does not add up if whole sample of married and non-married was analysed together. It also makes sense to only analyse married men as they were the only ones who were able to reproduce. Not explicit in papers which variables were transformed to normality.
Apicella, 2014	We excluded DVs >1 spouse in lifetime (considered redundant) and number of offspring born (effect size is the same for another variable but *p*-value differs - N is not stated, suggesting that either the effect size is not correct or N is considerably smaller). N for some analyses is not given, we assumed it was 51 as given for one of the analyses. They classified predictor as strength, but we re-coded it as a composite measure of muscle mass and handgrip strength as that is what it was (predictor was therefore not classified as either muscle mass or strength for moderation analyses). Some relationships reported in other papers as well: Smith, Olkhov, Puts & Apicella (2017) reported muscle mass/strength – reproductive success and offspring number; we kept results from this paper as it controlled for age, with the exclusion of offspring number for reason given above.
Apicella et al., 2007	Some relationships reported in other papers as well: Smith, Olkhov, Puts & Apicella (2017) reported f0 - reproductive success and offspring number; we kept results from this paper as it controlled for age.
Arnocky et al., 2018	Paper also included fWHR-lower - SOI-R and lower face/face height - SOI-R but effect sizes not reported separately for men and women so not included.
Atkinson et al., 2012	Paper included both DVs number of living children and genetic vector; the latter calculated as 1*(number of living children) + ½*(number of living grandchildren). Considered redundant to include both, and to be consistent with other measures, we included number of living children.
Boothroyd et al., 2017	Agta sample: photographs taken from front or ¾ degree angle were coded as not frontal photographs.
Charles & Alexander, 2011	We excluded SOI (Clark, 2004) as it is redundant to SOI and SOI is the commonly used measure. Sample assumed to be non-fathers.
Falcon, 2016	We used average 2D:4D rather than R2D:4D/L2D:4D due to bigger sample size and we could not rule out the possibility of overlapping samples.
Farrelly et al., 2015	All participants were heterosexual (information provided by author). Author re-ran analyses based on whole sample and provided results which we used, so the results do not exactly match results reported in paper (in the paper the authors had omitted a few participants due to incomplete relationship information).

Frederick & Jenkins, 2015	The paper also included dichotomous variables: more than 5 sex partners and more than 14 sex partners. We did not include those (considered redundant) and instead only used the continuous variable number of sex partners.
Gallup et al., 2007	The paper included both SHR circumference and breadth; we only included circumference to keep it consistent with other results.
Genovese, 2008	In the paper, the relationship between HGS and height was reported, but the paper did not include N or information about whether age was controlled for, so first author re-ran analyses. For mesomorphy - offspring number, we assumed N=181 as first author could only find (reliable) data on offspring number for 181 participants in the primary data source.
Gettler et al., 2019	Fertility in The Philippines has now dropped below 3.0 children/woman but was above in 2009 when data was collected (according to https://data.worldbank.org/indicator/SP.DYN.TFRT.IN?locations=PH), therefore coded as a high fertility sample.
Hartl et al., 1982	HGS - offspring number and height - offspring number were also analyzed and reported by Genovese (2008), but that paper did not include N and it was not clear whether age had been controlled for, so first author re-ran analyses. Participants with clearly incomplete or inaccurate family histories were excluded. In cases where family history was clear up to a certain point, or the participant had died, their age at that point was used. Thus, some of these relationships are reported as non-peer reviewed and some as peer-reviewed (the latter in the case where the relationship was reported by Genovese). We did not include general strength, as it was assessed subjectively.
Hoppler, Walther et al., 2018	Ninety-seven percent of sample was white (mentioned in other paper on the same sample).
Hönekopp et al., 2007	Same sample as in Hönekopp et al., 2006, who reported that 80% of the sample were students.
Kirchengast, 2000	Judged to be the same sample as Winkler and Kirchengast, 1994.
Kirchengast & Winkler, 1995	Mean number of children in sample: 1.1 in Rundu and 1.8 in rural areas; however, age range of sample was 18-39 and mean age = 26 so it is young sample with non-completed reproductive histories, therefore coded as a high fertility sample.
Klimas et al., 2019	Ninety-seven percent of sample was white (mentioned in other paper on same sample). Only included men without sexual dysfunction. Excluded participants who had had bleeding or injuries in the mouth in the last few days before testing (clear from other paper on same sample), so we considered blood contamination of saliva sample controlled for.
Little et al., 1989	Population mean and S.D. for height stated in another paper (by the same author); we used that to convert effect sizes.
Loehr & O'Hara, 2013	We assumed it was primarily a white sample.

Longman et al., 2018	Given that this was a young British student sample, we assumed that they were all non-fathers. Baseline testosterone was, in a sense, anticipatory, but we included this paper since effect sizes in this study did not differ substantially from effect sizes in other studies.
Lukaszewski et al., 2014	The first author ran analyses on openly available data. Some extreme outliers in one of the samples (chest strength around ~2, which seems incorrect). However, it made no difference to the results whether they were removed or kept, so we kept them in.
Marczak et al., 2018	2D:4D was measured directly as well as from digital photos. Not explicit whether that meant that measurement method varied between participants. As we could not be sure that all participants had been measured directly, we coded this as hand scans.
Mosing et al., 2015	Published paper but author submitted results to us. Twin sample; only unrelated individuals included in this sample. We coded it as a heterosexual sample (in the paper, gay participants were excluded), and as a predominantly white sample.
Mueller & Mazur, 1997	N was not explicit in the results, but they stated that 337 participants replied so we assumed N=337. Sample was born 1923-1929 so should have completed most of their reproduction by 1965 when the fertility rate dropped below 3 in the U.S.; therefore coded as a high fertility, industrialised sample.
Nagelkerke et al., 2006	Author sent us raw data. For age at first sexual intercourse, we set cut off at 12 (there were data points <12) as we deemed it unlikely that participants had had sexual onset prior to puberty.
Nettle, 2002	This sample's parents were analyzed in Krzyzanowska et al. (2015) but as the parents reproduced separately and at a different time point compared to this sample, we considered them to be separate samples.
Pawlowski et al., 2008	We assumed this sample was heterosexual.
Pawlowski et al., 2000	Author sent us the results. Coded as a high fertility, industrialised sample.
Polo et al., 2019	The paper also included skeletal muscle mass but this measure was extremely highly correlated with upper-body fat free mass (FFM: $r = .96$, n = 206, $p < .001$) so it was considered redundant to include both predictors, and to be consistent with other studies we kept upper-body FFM. Sample consisted of heterosexual students and non-students: we did not know the proportion of students versus non-students, so we coded it as a non-student sample. Results were considered published as these relationships were reported in paper, although the paper included a type of analyses that we could not use and the authors therefore submitted r.
Puts et al., 2015	We assumed the sample were non-fathers. For one sample, testosterone was sampled by saliva between 9AM-1.30PM; we coded that as AM testosterone.

Puts et al., 2006	We judged the sample to be the same one as Putz et al. (2004) and Hodges-Simeon et al. (2011).
Putz et al., 2004	We judged the sample in study 1 to be the same one as Puts et al. (2006) and Hodges-Simeon et al. (2011).
Rahman et al., 2005	Proportion of students vs non-students not clear, so coded as unknown low fertility sample.
Rosenfield et al., 2020	Coded as heterosexual sample as all participants had been married (heterosexually) at some point.
Scott & Bajema, 1982	The paper reported both zero-order correlations and partial correlations controlling for ethnicity. We used zero-order correlations because the sample was from the same group, even if their ethnicities differed. Sample was born 1912-1918 and should therefore have largely completed reproduction by 1965 when the fertility rate dropped below 3 in the U.S.; therefore coded as a high fertility, industrialised sample.
Sim & Chun, 2016	We judged the sample to be the same one as Sim (2013). SHR also reported in Sim (2013); we therefore excluded that paper.
Smith et al., 2017	Same sample/analyses as reported in Apicella et al. (2007) and Apicella (2014); kept those papers as those analyses controlled for age, with the exception of muscle mass/strength - offspring number (also given in Apicella, 2014), as in the latter, N was not explicit.
Steiner et al., 2011	Paper also included SOI and extrapair sexual interest (EPSI), but SOI and EPSI were measured after viewing a video which, for some participants, had sexual content, and those variables were therefore measured after manipulation so we did not include them. Sample consisted of 90% exclusively heterosexual participants, and 10% heterosexual but incidentally gay participants; coded as a heterosexual sample.
Stern et al., 2020	Information about ethnicity not available in paper; however, it was given in other paper on same sample (Kandrik et al., 2016) in which it was reported that 91% of a subsample was white. This sample was therefore coded as predominantly white.
Strong et al., 2014	We excluded 2D:4D following recommendation from author, due to potential measurement issues.
Suire et al., 2018	Baseline recording consisted of just a short utterance repeated after the experimenter, and author therefore suggested to use one of the other recordings (courtship and competition) instead. Results did not show any substantial difference between the recordings, however, so we used baseline to be consistent with other studies. Data set contained virgins: current age set as age at first sexual intercourse, as is commonly done.
Tao & Yin, 2016	Offspring number: S.D. of the mean for the whole sample not given but varied between 1.1-1.3 for each of the three samples so we used that to convert effect sizes (which value was used did not affect the results in any

	case). We used whole sample rather than the three sub-samples, coded as low fertility.
van Anders et al., 2007	Sampled from a monogamous society, but some of the samples were polyamorous, therefore coded as non-monogamous.
van Dongen & Sprengers, 2012	Sample not specified but we assumed it was from a low fertility, monogamous population.
Varella et al., 2014	Assumed N=80 for all non-significant relationships where N was not specified, as stated elsewhere in the paper. We used results for the two samples combined, not the sub-samples.
von Rueden et al., 2010	DVs not normally distributed but could be transformed to near-normality.
Walther et al., 2016	Ninety-seven percent of sample was white (mentioned in other paper on same sample).
Walther et al., 2017b	Ninety-seven percent of sample was white (mentioned in other paper on same sample).
Walther et al., 2017c	Ninety-seven percent of sample was white (mentioned in other paper on same sample).
Weeden & Sabini, 2007	Paper also included Sociosexuality measure; however, as we did not know the response scale or direction of responses, we excluded it.
Winkler & Kirchengast, 1994	Judged to be the same sample as Kirchengast, 2000.

Note. 2D:4D = 2nd to 4th finger (digit) ratio; f0 = fundamental frequency, i.e., voice pitch; fWHR = facial width-to-height ratio; HGS = handgrip strength; SHR = shoulder-to-hip ratio; SOI = Sociosexual Orientation Inventory; SOI-R = the Revised Sociosexual Orientation Inventory.

Appendix C

All studies included in the meta-analysis

Authors	Year	Predictor	Outcome	Sample	Sample location	Low or high fert.	*N*
Alvergne et al.	2009	T	REP	Rural villagers	Senegal	High	53
Apicella	2014	Body masc	MAT, REP, OM	Hadza	Tanzania	High	51
Apicella et al.	2007	Body masc, voice pitch, height	REP, OM	Hadza	Tanzania	High	44-52
Arnocky et al.	2018	Facial masc	MAT	Students	Canada	Low	135
Aronoff et al.	2017	T	MAT	Students	US	Low	99
Atkinson	2012	Body masc	MAT	Students	US	Low	66
Atkinson et al.	2012	Body masc, voice pitch, height	REP	Himba (Ovahimba)	Namibia	High	36
Bogaert & Fisher	1995	T	MAT	Students	Canada	Low	195-196
Booth et al.	1999	T	MAT	Army veterans and non-veterans	US	Low	4393
Boothroyd et al.	2008	Facial masc	MAT	Students	UK	Low	18-19
Boothroyd et al.	2011	Facial masc	MAT	Students	UK	Low	36
Boothroyd et al.	2017	Facial masc	REP, OM	Agta	Philippines	High	65
		Facial masc	MAT, REP, OM	Maya	Belize	High	23-35
Charles & Alexander	2011	2D:4D, T	MAT	Students	US	Low	25-42
Chaudhary et al.	2015	Body masc, height	MAT, REP, OM	Mbendjele BaYaka	Democratic Republic of the Congo	High	55-73
Edelstein et al.	2011	T	MAT	Students	US	Low	134
Falcon	2016	2D:4D	MAT	Students	US	Low	137
Farrelly et al.	2015	T	MAT	Students	UK	Low	75-78
Frederick	2010	Body masc, 2D:4D, height	MAT	Students	US	Low	61

Frederick & Haselton	2007	Body masc	MAT	Students	US	Low	56-121
Frederick & Jenkins	2015	Height	MAT	Online	Worldwide	Low	28759-31418
Gallup et al.	2007	Body masc, 2D:4D	MAT	Students	US	Low	71-75
Genovese	2008	Body masc	REP	Former teenage delinquents	US	High	181
Gettler et al.	2019	T	MAT	Cebu Longitudinal Health and Nutrition Survey	Philippines	High	288
Gildner	2018	Body masc, 2D:4D, height	REP	Shuar Health and Life History Project	Ecuador	High	48
Gómez-Valdés et al.	2013	Facial masc	REP	Hallstatt skulls	Austria	High	179
Hartl et al.	1982	Body masc, height	MAT, REP	Former teenage delinquents	US	High	180-185
Hill et al.	2013	Facial masc, body masc, voice pitch, height	MAT	Students	US	Low	63
Hoppler, Walther et al.	2018	T	REP	Men’s health 40+ study	Switzerland	Low	268
Hughes & Gallup Jr.	2003	Body masc	MAT	Students	US	Low	50-59
Hönekopp et al.	2006	2D:4D, height	MAT	Students and non-students	Germany	Low	79-99
Hönekopp et al.	2007	Facial masc, body masc, height, T	MAT	Students and non-students	Germany	Low	77
Kirchengast	2000	Height	REP, OM	!Kung San	Namibia	High	103
Kirchengast & Winkler	1995	Height	REP, OM	Urban and rural Kavango people	Namibia	High	59-78
Klimas et al.	2019	T	MAT	Men’s health 40+ study	Switzerland	Low	159
Klimek et al.	2014	2D:4D, height	REP	Mogielica Human	Poland	High	238

				Ecology Study Site			
Kordsmeyer et al.	2018	Body masc, voice pitch, height, T	MAT	Students and non-students	Germany	Low	103-164
Kordsmeyer & Penke	2017	2D:4D, height	MAT	Students and non-students	Germany	Low	141
Krzyzanowska et al.	2015	Height	REP	National Child Development Study	UK	Low	6535
Kurzban & Weeden	2005	Height	MAT, REP	Speed daters	US	Low	1503-1501
Lassek & Gaulin	2009	Body masc, height	MAT	NHANES III	US	Low	4167-5159
Little et al.	1989	Height	REP, OM	Rural; growth stunted	Mexico	High	103
Loehr & O'Hara	2013	Facial masc	REP	WWII soldiers	Finland	High	795
Longman et al.	2018	T	MAT	Students	UK	Low	38
Luevano et al.	2018	Facial masc, height	MAT	Students	US	Low	35-66
Lukaszewski et al.	2014	Body masc	MAT	Students	US	Low	48-174
Maestripieri et al.	2014	T	MAT	Students	US	Low	41-61
Manning & Fink	2008	2D:4D	MAT, REP	Online	Worldwide	Low	26872-83681
Manning et al.	2003	2D:4D	REP	Community	England	Low	189
		2D:4D	REP	Sugali and Yanadi tribal groups	India	High	80
		2D:4D	REP	Zulus from townships near Durban	South Africa	High	66
Marczak et al.	2018	2D:4D	REP	Yali	Indonesia	High	47
McIntyre et al.	2006	T	MAT	Students	US	Low	68-81
Međedović & Bulut	2019	Height	MAT	Students	Serbia	Low	39
Mosing et al.	2015	Height	MAT, REP	Study of Twin Adults: Genes and Environment	Sweden	Low	2310-2549
Muller & Mazur	1997	Facial masc	REP	West Point class of 1950	US	High	337

Nagelkerke et al.	2006	Height	MAT	NHANES 99–00	US	Low	798-809
Nettle	2002	Height	REP	National Child Development Study	UK	Low	4474
Pawlowski et al.	2008	Height	REP	Rural	Poland	High	46
Pawlowski et al.	2000	Height	REP	Urban and rural	Poland	High	3201
Peters et al.	2008	Facial masc, body masc, T	MAT	Students	Australia	Low	100-113
Pollet et al.	2011	T	MAT	National Social Life, Health, and Aging Project	US	Low	749
Polo et al.	2019	Facial masc, body masc, height	MAT	Students and non-students	Chile	Low	198-206
Price et al.	2013	Body masc, height	MAT	Mainly students	UK	Low	55
Prokop & Fedor	2011	Height	REP	Friends and family of students	Slovakia	Low	499
Prokop & Fedor	2013	Height	MAT	Students	Slovakia	Low	105-150
Puts et al.	2006	Voice pitch	MAT	Students	US	Low	103
Puts et al.	2015	T	MAT	Students	US	Low	59-61
Putz et al.	2004	2D:4D	MAT	Students	US	Low	207-219
Rahman et al.	2005	2D:4D, height	MAT	Students and non-students	UK	Low	78-150
Rhodes et al.	2005	Facial masc, body masc, height	MAT	Mainly students	Australia	Low	142-166
Rosenfield et al.	2020	Body masc, voice pitch, height	MAT, REP, OM	Tsimané	Bolivia	High	55-62
Schwarz et al.	2011	2D:4D	MAT	Students	Germany	Low	52-89
Scott & Bajema	1982	Height	REP	Third Harvard Growth Study	US	High	606

Shoup & Gallup	2008	Body masc, 2D:4D	MAT	Students	US	Low	28-38
Sim & Chun	2016	Body masc, 2D:4D	MAT	Students	US	Low	90
Simmons & Roney	2011	Body masc, T	MAT	Students	US	Low	138
Smith et al.	2017	Body masc	REP	Hadza	Tanzania	High	51
Sneade & Furnham	2016	Body masc	MAT	Students	UK	Low	145
Sorokowski et al.	2013	Height	REP, OM	Yali	Indonesia	High	49-52
Steiner et al.	2011	2D:4D, T	REP	Students and non-students	US	Low	30
Stern et al.	2020	T	MAT	Students	UK	Low	61
Strong et al.	2014	Body masc	MAT	Students	US	Low	31
Strong & Luevano	2014	Body masc, 2D:4D, height	MAT	Students	US	Low	51-66
Subramanian et al.	2009	Height	OM	2005-2006 National Family Health Survey	India	Low	21120
Suire et al.	2018	Voice pitch	MAT	Mainly students	France	Low	57-58
Tao & Yin	2016	Height	REP	The Panel Study of Family Dynamics	Taiwan	Low	1409
van Anders et al.	2007	T	MAT	Non-students	US	Low	31
Van Dongen & Sprengers	2012	Facial masc, body masc, 2D:4D	MAT	Not specified	Not specified	Low	52
Varella et al.	2014	Body masc, 2D:4D, height	MAT	Students	Brazil, Czech Republic	Low	69-80
von Rueden et al.	2011	Body masc, height	REP, OM	Tsimané	Bolivia	High	162-197
Voracek et al.	2010	2D:4D, height	REP	Firefighters	Austria	Low	134
Walther et al.	2016	Body masc	REP	Men’s health 40+ study	Switzerland	Low	271
Walther et al.	2017a	Body masc	MAT	Men’s health 40+ study	Switzerland	Low	226
Walther et al.	2017b	Height	REP	Men’s health 40+ study	Switzerland	Low	271

Walther et al.	2017c	Height	MAT	Men’s health 40+ study	Switzerland	Low	226
Waynforth	1998	2D:4D, height	MAT, REP, OM	Villagers	Belize	High	35-56
Weeden & Sabini	2007	Body masc, 2D:4D, height	MAT	Students	US	Low	188-212
Winkler & Kirchengast	1994	Height	REP, OM	!Kung San	Namibia	High	31-114

Note. Fert = fertility, masc = masculinity, MAT = mating, NHANES = National Health and Nutrition Examination Survey, OM = offspring mortality, REP = reproduction, T = testosterone.

Appendix D.1

General moderators: for all predictors

Moderator	Description
Domain type	Mating vs reproductive domain.
Mating measure type	Mating attitudes (e.g., preferences for short-term relationships/casual sex) vs mating behaviours (e.g., number of sexual partners, age at first sexual intercourse)
Reproductive measure type	Fertility (i.e., number of children/grandchildren, age at the birth of the first child) vs reproductive success (i.e., number of surviving children/grandchildren).
Sample type	Low fertility samples (i.e., <3.0 children/woman within sample/population at the time of sampling) vs high fertility samples. The latter is considered to correspond to naturally fertile populations.
Low fertility samples	Predominantly student samples (i.e., ≥ 50% students) vs mixed/non-student/unknown samples.
High fertility samples	Traditional vs industrialised samples.
Ethnicity	Predominantly white (i.e., ≥ 75% of sample) vs mixed/non-white /unknown.
Marriage system	Monogamy vs non-monogamy/unknown.
Publication status	Published vs non-published results. We favoured publication status of the relevant *results* rather than of the *paper*, since we retrieved many of our effects from published studies where the key relationship had not been analyzed/was not the focus of the paper.
Peer-review	Peer-reviewed vs not peer-reviewed study.
Sexual orientation	Heterosexual sample vs gay/mixed/unknown sexual orientation.
Normality-transformed variables	Non-transformed vs transformed variables, i.e., whether skewed variables (skew is very common for some of the variables, such as number of sexual partners) had been e.g., log-transformed to normality.
Converted effect size	Non-converted vs converted effect sizes, i.e., whether effect size was given as Pearson's *r* or whether we had used a formula to convert it. The latter results in an estimate of *r*.
Age control	Age controlled for in analyses vs not controlled for. We considered age an essential control for all analyses except *i*. where all participants belonged to the same age group, *ii*. for the variables sexual onset/reproductive onset, and *iii*. mating attitudes.

Non-relevant controls	No non-relevant vs non-relevant controls included in the analyses. For example, analyses with several non-relevant predictors may produce weaker associations compared to e.g., bivariate correlations with just the relevant predictor and outcome variables.

Note. We were constrained by information made available in papers. The levels of moderators should therefore be considered to reflect where we knew for certain that a moderator e.g., had been controlled for vs where we could not be certain. For several of our potential moderators, such as the moderators we had selected for voice pitch, not enough papers mentioned having controlled for them and we were therefore unable to analyse those moderators. Additionally, we often did not have enough observations on each level of the moderator variable to be able to run those analyses. This lack of power also prevented us from analysing combined effects of several moderators; we therefore analyzed moderators one by one. Moderators were coded into two levels wherever possible, as otherwise we would often have had too few observations/level to be able to run the analysis. It should also be noted that some moderators are likely confounded; for example, non-monogamous populations are almost always high fertility, traditional populations.

Appendix D.2

Facial masculinity moderators

Moderator	Description
Measurement type	Objectively measured masculinity (using geometric morphometric analyses) vs observer-rated masculinity vs fWHR (i.e., facial width-to-height ratio).
Standardization of photographs	Photographs taken under standardized vs not standardized/semi-standardized/unknown conditions.
Angle of photographs	Front-facing vs not front-facing/unknown angle of photographs.
Masked photographs	Masked vs not masked photographs/unknown. Only coded for rated facial masculinity.
Adiposity	Adiposity/body mass index (BMI) controlled for vs not controlled for/unknown. Only coded for rated facial masculinity.
Colour vs black & white photographs	Colour vs black & white photographs/unknown. Only coded for rated facial masculinity.
Facial expression	Neutral vs smiling/mixed/unknown facial expressions. Only coded for rated facial masculinity.
Facial hair	Clean-shaven vs not clean-shaven/mixed/unknown. Only coded for rated facial masculinity.

Note. Moderators were coded into two levels wherever possible, as otherwise we would often have had too few observations/level to be able to run the analysis.

Appendix D.3

Body masculinity moderators

Moderator	Description
Number of measurements	Only coded for measured body masculinity. Typically referred to repeat measurements but in some cases, different measurements were used.
Adiposity	Adiposity/BMI controlled for vs not controlled for/unknown. Only coded for rated body masculinity.
Measurement type	Measured vs observer- or own-rated body masculinity. Measured body masculinity included e.g. strength, circumference of shoulder-to-hip ratio, and bioelectrical measurement of fat-free mass.
Body masculinity type	Strength vs body shape vs muscle mass. Strength was typically assessed through measured handgrip strength. Body shape included body measurements (see measurement type above) and rated body masculinity. Muscle mass was measured (see above) or rated.

Note. Moderators were coded into two levels wherever possible, as otherwise we would often have had too few observations/level to be able to run the analysis.

Appendix D.4

2D:4D moderators

Moderator	Description
Measurement type	Measured directly vs measured from hand scans/photographs vs self-reported vs unknown.
Number of measurements	Only coded for experimenter-measured (directly or from hand scans).
Finger injuries	Controlled for vs not controlled for.
Left vs right	Left vs right hand 2D:4D. 2D:4D dimorphism is typically claimed to be more pronounced in the right hand (Hönekopp et al., 2006).

Note. Moderators were coded into two levels wherever possible, as otherwise we would often have had too few observations/level to be able to run the analysis.

Appendix D.5

Voice pitch moderators

Moderator	Description
Sex of experimenter	Female vs male vs unknown.
Illness	Illnesses (colds etc. that could influence voice pitch) controlled for/excluded vs not.
Smoker	Smoking participants controlled for/excluded vs not.
Condition	Baseline vs courtship cs competitive type of recording.

Note. Based on information available in the papers, none of these potential moderators had been controlled for in any of the studies.

Appendix D.6

Height moderators

Moderator	Description
Measurement type	Experimenter-measured vs self-reported vs unknown.
Number of measurements	Only coded for experimenter-measured height.

Note. Moderators were coded into two levels wherever possible, as otherwise we would often have had too few observations/level to be able to run the analysis.

Appendix D.7

Testosterone levels moderators

Moderator	Description
How assayed	Assayed from blood vs saliva vs unknown.
Time of day	Assayed in the AM vs PM vs unknown.
Blood contamination	Checked for vs not checked for/unknown. Only coded for saliva assayed T levels.
Fatherhood	Non-fathers vs fathers vs mixed/unknown
Relationship status	Married/in committed relationship vs single vs mixed/unknown.

Note. Moderators were coded into two levels wherever possible, as otherwise we would often have had too few observations/level to be able to run the analysis.

Appendix E.1

Facial masculinity: moderation analyses. The intercept shows the 'simple effect' for the reference category (specified) and the moderator effect shows the change in effect size for that category relative to the reference category. Moderators are bolded if significant after controlling for multiple comparisons, as indicated by computation of q-values. The full list of q-values can be found in Supplementary File 7.

Mating vs reproductive domain

		B	SE	[95% CI]	*z*	*p*
Domain type	Intercept (mating domain)	0.088	0.042	0.007, 0.169	2.120	.034
	Reproductive domain	0.006	0.072	-0.135, 0.148	0.089	.929

Mating domain (MAT), mating behaviours & mating attitudes

		B	SE	[95% CI]	*z*	*p*
MAT measure type	Intercept (MAT behaviours)	0.047	0.041	-0.033, 0.128	1.150	.250
	MAT attitudes	0.038	0.061	-0.082, 0.159	0.622	.534
Ethnicity	Intercept (predominantly white)	0.114	0.057	0.003, 0.225	2.007	.045
	Mixed/other/unknown	-0.075	0.084	-0.241, 0.090	-0.892	.373
Publication status	Intercept (published results)	0.046	0.058	-0.067, 0.159	0.801	.423
	Non-published results	0.092	0.093	-0.090, 0.273	0.990	.322
Publication status: MAT behaviours	Intercept (published results)	0.016	0.056	-0.094, 0.126	0.281	.778
	Non-published results	0.023	0.098	-0.169, 0.215	0.233	.816
Sexual orientation	Intercept (heterosexual sample)	0.044	0.055	-0.063, 0.151	0.806	.420
	Gay/mixed/unknown	0.105	0.092	-0.075, 0.285	1.145	.252
Age control	Intercept (age controlled for)	0.065	0.060	-0.052, 0.181	1.086	.277
	Age not controlled for	0.038	0.098	-0.154, 0.230	0.387	.699
Age control: MAT behaviours	Intercept (age controlled for)	0.017	0.052	-0.086, 0.119	0.318	.750

	Age not controlled for	0.024	0.100	-0.172, 0.220	0.241	.810
Measurement type	Intercept (measured)	0.105	0.058	-0.008, 0.219	1.825	.068
	Rated	-0.009	0.079	-0.162, 0.146	-0.108	.914
	fWHR: *s* =2					

Other moderators with too few *k*/*s*:

Sample type: low vs high fertility; Low fertility samples: predominantly students vs non-students; High fertility samples: traditional vs industrialised; Marriage system: monogamy vs non-monogamy; Peer-reviewed vs not peer-reviewed; Normality-transformed variables; Converted effect size; Non-relevant controls, Standardization of photographs; Angle of photographs; Masked photographs; Adiposity; Colour vs black & white photographs; Facial expression; Facial hair.

Reproductive domain (REP), fertility & reproductive success

Moderators with too few *k*/*s*:

REP measure type: reproductive success vs fertility; Sample type: low vs high fertility; Low fertility samples: predominantly students vs non-students; High fertility samples: traditional vs industrialised; Ethnicity: predominantly white vs not; Marriage system: monogamy vs non-monogamy; Publication status: published results; Peer-reviewed; Sexual orientation: heterosexual sample vs gay/mixed/unknown; Normality-transformed variables; Converted effect size; Age controlled for; Non-relevant controls, Measurement type: measured vs rated vs fWHR; Standardization of photographs; Angle of photographs; Masked photographs; Adiposity; Colour vs black & white photographs; Facial expression; Facial hair.

Note. *k* = number of observations, MAT = mating, REP = reproductive, *s* = number of samples. Moderation analyses were only run where each level of the moderator included observations from at least two studies and three independent samples. Analyses were run on the mating measures mating behaviours and mating attitudes, and the reproductive measures fertility and reproductive success when there were enough observations to do so.

Appendix E.2

Body masculinity: moderation analyses. The intercept shows the 'simple effect' for the reference category (specified) and the moderator effect shows the change in effect size for that category relative to the reference category. Moderators are bolded if significant after controlling for multiple comparisons, as indicated by computation of q-values. The full list of q-values can be found in Supplementary File 7.

Mating vs reproductive domain						
		B	SE	[95% CI]	*z*	*p*
Domain type	Intercept (mating domain)	0.132	0.021	0.092, 0.173	6.414	<.001
	Reproductive domain	0.019	0.042	-0.064, 0.102	0.439	.661

Mating domain (MAT), mating behaviours & mating attitudes						
		B	SE	[95% CI]	*z*	*p*
MAT measure type	Intercept (MAT behaviours)	0.139	0.022	0.096, 0.181	6.382	<.001
	MAT attitudes	-0.024	0.031	-0.085, 0.037	-0.773	.440
Sample type	Intercept (low fertility)	0.136	0.023	0.091, 0.181	5.878	<.001
	High fertility	-0.024	0.078	-0.177, 0.129	-0.307	.759
Sample type: MAT behaviours	Intercept (low fertility)	0.147	0.024	0.099, 0.194	6.019	<.001
	High fertility	-0.035	0.078	-0.188, 0.119	-0.443	.658
Low fertility sample	Intercept (predominantly students)	0.157	0.023	0.111, 0.203	6.699	<.001
	Non-students/ mixed/unknown	-0.118	0.051	-0.218, -0.019	-2.336	.020
Low fertility sample: MAT behaviours	Intercept (predominantly students)	0.172	0.024	0.125, 0.218	7.285	<.001
	Non-students/ mixed/unknown	**-0.128**	**0.049**	**-0.224, -0.033**	**-2.632**	**.009**
Ethnicity	Intercept (predominantly white)	0.116	0.044	0.030, 0.203	2.643	.008

	Mixed/other/unknown	0.024	0.051	-0.076, 0.124	0.464	.643
Ethnicity: MAT behaviours	Intercept (predominantly white)	0.116	0.044	0.031, 0.202	2.661	.008
	Mixed/other/unknown	0.038	0.052	-0.064, 0.139	0.728	.467
Marriage system	Intercept (monogamy)	0.139	0.022	0.095, 0.182	6.230	<.001
	Non-monogamy	-0.095	0.096	-0.283, 0.093	-0.991	.322
Marriage system: MAT behaviours	Intercept (monogamy)	0.149	0.023	0.104, 0.195	6.422	<.001
	Non-monogamy	-0.106	0.096	-0.294, 0.082	-1.104	.270
Publication status	Intercept (published results)	0.167	0.026	0.117, 0.218	6.470	<.001
	Non-published results	-0.086	0.039	-0.163, -0.009	-2.181	.029
Publication status: MAT attitudes	Intercept (published results)	0.099	0.044	0.013, 0.184	2.251	.024
	Non-published results	-0.079	0.081	-0.237, 0.079	-0.975	.330
Publication status: MAT behaviours	Intercept (published results)	0.177	0.028	0.123, 0.231	6.402	<.001
	Non-published results	-0.087	0.044	-0.172, -0.001	-1.978	.048
Peer-review	Intercept (peer-reviewed)	0.136	0.024	0.088, 0.184	5.576	<.001
	Not peer-reviewed	-0.012	0.058	-0.126, 0.102	-0.204	.838
Peer-review: MAT behaviours	Intercept (peer-reviewed)	0.142	0.025	0.093, 0.192	5.636	<.001
	Not peer-reviewed	0.005	0.063	-0.118, 0.128	0.085	.933
Sexual orientation	Intercept (heterosexual sample)	0.177	0.030	0.118, 0.235	5.948	<.001

	Gay/mixed/unknown	-0.085	0.041	-0.165, -0.006	-2.098	.036
Sexual orientation: MAT attitudes	Intercept (heterosexual sample)	0.045	0.057	-0.067, 0.157	0.781	.435
	Gay/mixed/unknown	0.062	0.077	-0.089, 0.212	0.804	.421
Sexual orientation: MAT behaviours	Intercept (heterosexual sample)	0.188	0.031	0.127, 0.249	6.069	<.001
	Gay/mixed/unknown	-0.088	0.042	-0.171, -0.006	-2.091	.037
Normality-transformed variables	Intercept (non-transformed variables)	0.137	0.025	0.088, 0.185	5.523	<.001
	Transformed variables	0.038	0.047	-0.054, 0.129	0.810	.418
Normality-transformed variables: MAT behaviours	Intercept (non-transformed variables)	0.049	0.037	-0.024, 0.122	1.321	.186
	Transformed variables	**0.165**	**0.054**	**0.060, 0.270**	**3.091**	**.002**
Converted effect size	Intercept (not converted)	0.144	0.026	0.093, 0.194	5.587	<.001
	Converted	0.003	0.063	-0.121, 0.128	0.053	.958
Converted effect size: MAT behaviours	Intercept (not converted)	0.152	0.028	0.098, 0.206	5.508	<.001
	Converted	-0.005	0.065	-0.133, 0.123	-0.076	.940
Age control	Intercept (age controlled for)	0.098	0.031	0.037, 0.158	3.147	.002
	Age not controlled for	**0.103**	**0.042**	**0.020, 0.186**	**2.441**	**.015**
Age control: MAT behaviours	Intercept (age controlled for)	0.107	0.033	0.043, 0.171	3.277	.001
	Age not controlled for	0.096	0.045	0.009, 0.183	2.153	.031

Number of measurements	Intercept (unknown number of measurements)	0.076	0.041	-0.005, 0.156	1.851	.064
	2 measurements	0.046	0.047	-0.046, 0.137	0.974	.330
	3 measurements	0.126	0.062	0.004, 0.247	2.025	.043
	1 measurement: *s* = 1					
Number of measurements: MAT attitudes	Intercept (unknown number of measurements)	0.117	0.054	0.010, 0.223	2.144	.032
	2 measurements	-0.070	0.071	-0.210, 0.069	-0.987	.324
	1 measurement: *s* = 1					
	3 measurements: *s* = 0					
Number of measurements: MAT behaviours	Intercept (unknown number of measurements)	0.061	0.048	-0.034, 0.155	1.259	.208
	2 measurements	0.093	0.059	-0.023, 0.210	1.567	.117
	3 measurements	0.151	0.070	0.015, 0.288	2.169	.030
	1 measurement: *s* = 1					
Measurement type	Intercept (measured)	0.081	0.040	0.002, 0.159	2.011	.044
	Rated	**0.177**	**0.066**	**0.048, 0.306**	**2.695**	**.007**
Measurement type: MAT behaviours	Intercept (measured)	0.087	0.041	0.007, 0.167	2.121	.034
	Rated	**0.174**	**0.066**	**0.044, 0.303**	**2.630**	**.009**
Body masculinity type	Intercept (strength)	0.187	0.031	0.126, 0.248	5.974	<.001
	Body shape	**-0.099**	**0.034**	**-0.165, -0.033**	**-2.945**	**.003**
	Muscle mass	-0.108	0.071	-0.247, 0.031	-1.529	.126
Body masculinity type: MAT behaviours	Intercept (strength)	0.205	0.035	0.136, 0.274	5.8130	<.001

	Body shape	**-0.105**	**0.040**	**-0.184, -0.026**	**-2.615**	**.009**
	Muscle mass	-0.124	0.074	-0.269, 0.021	-1.676	.094
Other moderators with too few *k*/*s*: High fertility sample: traditional vs industrialised; Non-relevant controls, Adiposity.						
Reproductive domain (REP), fertility & reproductive success						
		B	SE	[95% CI]	*z*	*p*
REP measure type	Intercept (Reproductive success)	0.170	0.071	0.032, 0.309	2.417	.016
	Fertility	-0.034	0.081	-0.192, 0.125	-0.417	.677
Publication status	Intercept (published results)	0.222	0.065	0.095, 0.349	3.418	.001
	Non-published results	-0.107	0.075	-0.254, 0.040	-1.423	.155
Normality-transformed variables	Intercept (non-transformed variables)	0.136	0.046	0.045, 0.227	2.928	.003
	Transformed variables	0.022	0.071	-0.117, 0.161	0.311	.756
Normality-transformed variables: Fertility	Intercept (non-transformed variables)	0.095	0.042	0.012, 0.178	2.246	.025
	Transformed variables	0.078	0.063	-0.045, 0.200	1.239	.215
Converted effect size	Intercept (not converted)	0.089	0.034	0.023, 0.155	2.635	.008
	Converted	**0.143**	**0.059**	**0.028, 0.258**	**2.437**	**.015**
Non-relevant controls	Intercept (no non-relevant controls)	0.136	0.042	0.054, 0.219	3.232	.001
	Non-relevant controls	0.032	0.082	-0.128, 0.193	0.395	.693
Number of measurements	Intercept (unknown number of measurement)	0.117	0.055	0.010, 0.224	2.138	.033
	3 measurements	0.067	0.084	-0.098, 0.232	0.797	.426

	1 measurement: *s* = 0					
	2 measurements: *s* = 0					
Number of measurements: Reproductive success	Intercept (unknown number of measurement)	0.072	0.201	-0.323, 0.466	0.356	.722
	3 measurements	0.136	0.178	-0.213, 0.485	0.763	.446
	1 measurement: *s* = 0					
	2 measurements: *s* = 0					
Body masculinity type	Intercept (strength)	0.112	0.036	0.041, 0.183	3.108	.002
	Muscle mass	0.028	0.066	-0.101, 0.158	0.430	.667
	Body shape: *s* = 0					

Other moderators with too few *k*/*s*:

Sample type: low vs high fertility; Low fertility sample: students vs non-students; High fertility sample: traditional vs industrialised; Ethnicity: predominantly white vs not; Marriage system: monogamy vs non-monogamy; Peer-reviewed vs not peer-reviewed; Sexual orientation: heterosexual sample vs gay/mixed/unknown; Age controlled for; Adiposity; Measurement type: measured vs rated.

Note. *k* = number of observations, MAT = mating, REP = reproductive, *s* = number of samples. Moderation analyses were only run where each level of the moderator included observations from at least two studies and three independent samples. Analyses were run on the mating measures mating behaviours and mating attitudes, and the reproductive measures fertility and reproductive success when there were enough observations to do so.

Appendix E.3

2D:4D: moderation analyses. The intercept shows the 'simple effect' for the reference category (specified) and the moderator effect shows the change in effect size for that category relative to the reference category. Moderators are bolded if significant after controlling for multiple comparisons, as indicated by computation of q-values. The full list of q-values can be found in Supplementary File 7.

Mating vs reproductive domain

		B	SE	[95% CI]	*z*	*p*
Domain type	Intercept (mating domain)	0.050	0.018	0.015, 0.084	2.809	.005
	Reproductive domain	0.007	0.004	0.000, 0.014	1.997	.046

Mating domain (MAT), mating behaviours & mating attitudes

		B	SE	[95% CI]	*z*	*p*
MAT measure type	Intercept (MAT behaviours)	0.042	0.022	-0.002, 0.085	1.870	.062
	MAT attitudes	0.004	0.039	-0.072, 0.080	0.097	.923
Low fertility sample	Intercept (predominantly students)	0.036	0.022	-0.007, 0.078	1.646	.100
	Non-students/mixed/unknown	0.014	0.042	-0.069, 0.096	0.321	.749
Low fertility sample: MAT behaviours	Intercept (predominantly students)	0.033	0.023	-0.012, 0.077	1.419	.156
	Non-students/mixed/unknown	0.053	0.049	-0.042, 0.148	1.096	.273
Ethnicity	Intercept (predominantly white)	0.072	0.022	0.030, 0.115	3.333	.001
	Mixed/other/unknown	**-0.080**	**0.032**	**-0.143, -0.016**	**-2.462**	**.014**
Ethnicity: MAT attitudes	Intercept (predominantly white)	0.113	0.063	-0.010, 0.236	1.809	.071
	Mixed/other/unknown	-0.128	0.081	-0.287, 0.032	-1.572	.116
Ethnicity: MAT behaviours	Intercept (predominantly white)	0.085	0.026	0.034, 0.137	3.245	.001

	Mixed/other/unknown	**-0.088**	**0.036**	**-0.158, -0.017**	**-2.423**	**.015**
Publication status	Intercept (published results)	0.042	0.022	-0.000, 0.085	1.954	.051
	Non-published results	-0.020	0.040	-0.098, 0.059	-0.492	.623
Publication status: MAT behaviours	Intercept (published results)	0.046	0.027	-0.006, 0.098	1.719	.086
	Non-published results	-0.017	0.044	-0.103, 0.068	-0.396	.692
Peer-review	Intercept (peer-reviewed)	0.031	0.019	-0.007, 0.068	1.615	.106
	Not peer-reviewed	0.034	0.057	-0.078, 0.146	0.592	.554
Peer-review: MAT behaviours	Intercept (peer-reviewed)	0.029	0.021	-0.013, 0.071	1.349	.178
	Not peer-reviewed	0.070	0.063	-0.054, 0.193	1.103	.270
Sexual orientation	Intercept (heterosexual sample)	0.014	0.026	-0.038, 0.065	0.520	.603
	Gay/mixed/unknown	0.035	0.034	-0.031, 0.102	1.041	.298
Sexual orientation: MAT behaviours	Intercept (heterosexual sample)	0.024	0.033	-0.041, 0.089	0.731	.465
	Gay/mixed/unknown	0.024	0.042	-0.058, 0.105	0.566	.571
Normality-transformed variables	Intercept (non-transformed variables)	0.011	0.026	-0.039, 0.061	0.443	.658
	Transformed variables	**0.103**	**0.043**	**0.020, 0.187**	**2.416**	**.016**
Normality-transformed variables: MAT behaviours	Intercept (non-transformed variables)	0.009	0.026	-0.041, 0.060	0.368	.713

	Transformed variables	**0.102**	**0.040**	**0.024, 0.180**	**2.553**	**.011**
Age control	Intercept (age controlled for)	0.069	0.051	-0.031, 0.169	1.358	.175
	Age not controlled for	-0.033	0.057	-0.144, 0.078	-0.583	.560
Age control: MAT behaviours	Intercept (age controlled for)	0.077	0.056	-0.032, 0.186	1.387	.165
	Age not controlled for	-0.025	0.063	-0.148, 0.098	-0.401	.689
Number of measurements	Intercept (unknown number of measurements)	0.021	0.021	-0.020, 0.061	1.009	.313
	2 measurements	-0.023	0.030	-0.082, 0.035	-0.782	.434
	3 measurements	**0.102**	**0.037**	**0.030, 0.175**	**2.759**	**.006**
	1 measurement: *s* = 1					
Number of measurements: MAT behaviours	Intercept (unknown number of measurements)	0.021	0.021	-0.020, 0.061	1.009	.313
	2 measurements	-0.023	0.030	-0.082, 0.035	-0.782	.434
	3 measurements	**0.102**	**0.037**	**0.030, 0.175**	**2.759**	**.006**
	1 measurement: *s* = 1					
Measurement type	Intercept (directly)	0.004	0.032	-0.058, 0.066	0.128	.898
	Hand scans	0.091	0.042	0.008, 0.174	2.145	.032
	Unknown	0.005	0.048	-0.090, 0.098	0.093	.926
	Self-reported: *s* = 1					
Measurement type: MAT behaviours	Intercept (directly)	0.005	0.032	-0.059, 0.069	0.152	.879
	Hand scans	0.083	0.043	-0.002, 0.168	1.913	.056

	Unknown	0.000	0.049	-0.095, 0.096	0.006	.995
	Self-reported: *s* = 0					
Finger injuries	Intercept (finger injuries controlled for)	-0.002	0.046	-0.093, 0.088	-0.048	.962
	Finger injuries not controlled for	0.046	0.050	-0.053, 0.144	0.908	.364
Finger injuries: MAT behaviours	Intercept (finger injuries controlled for)	0.016	0.047	-0.077, 0.109	0.335	.738
	Finger injuries not controlled for	0.030	0.053	-0.074, 0.134	0.566	.572
Left vs right hand ratios	Intercept (right 2D:4D)	0.039	0.021	-0.002, 0.080	1.872	.061
	Left 2D:4D	-0.002	0.005	-0.013, 0.009	-0.363	.717
Left vs right hand ratios: MAT attitudes	Intercept (right 2D:4D)	0.006	0.066	-0.124, 0.136	0.093	.926
	Left 2D:4D	0.043	0.060	-0.076, 0.161	0.707	.479
Left vs right hand ratios: MAT behaviours	Intercept (right 2D:4D)	0.037	0.029	-0.021, 0.094	1.253	.210
	Left 2D:4D	0.014	0.032	-0.048, 0.077	0.449	.653

Other moderators with too few *k*/*s*:

Sample type: low vs high fertility; High fertility sample: traditional vs industrialised; Marriage system: monogamy vs non-monogamy; Converted effect size; Non-relevant controls.

Reproductive domain (REP), fertility & reproductive success

		B	SE	[95% CI]	*z*	*p*
REP measure type	Intercept (reproductive success)	0.170	0.066	0.042, 0.301	2.592	.010
	Fertility	-0.135	0.078	-0.289, 0.016	-1.751	.080
Sample type	Intercept (low fertility)	0.075	0.057	-0.037, 0.187	1.319	.187

	High fertility	0.002	0.072	-0.143, 0.140	-0.022	.983
High fertility sample	Intercept (traditional)	0.028	0.089	-0.145, 0.202	0.320	.749
	Industrialised	0.110	0.126	-0.137, 0.357	0.874	.382
Ethnicity	Intercept (predominantly white)	0.066	0.055	-0.042, 0.175	1.196	.232
	Mixed/other/unknown	0.016	0.072	-0.124, 0.157	0.227	.821
Marriage system	Intercept (monogamy)	0.054	0.042	-0.027, 0.136	1.304	.192
	Non-monogamy	0.116	0.089	-0.058, 0.290	1.306	.192
Publication status	Intercept (published results)	0.105	0.040	0.026, 0.184	2.601	.009
	Non-published results	-0.166	0.093	-0.348, 0.015	-1.795	.073
Sexual orientation	Intercept (heterosexual sample)	0.039	0.066	-0.091, 0.169	0.584	.560
	Gay/mixed/unknown	0.062	0.087	-0.109, 0.233	0.707	.480
Sexual orientation: Fertility	Intercept (heterosexual sample)	0.037	0.079	-0.119, 0.192	0.462	.644
	Gay/mixed/unknown	-0.019	0.125	-0.264, 0.225	-0.155	.877
Normality-transformed variables	Intercept (non-transformed variables)	0.031	0.036	-0.040, 0.101	0.851	.395
	Transformed variables	0.138	0.067	0.006, 0.269	2.054	.040
Age control	Intercept (age controlled for)	0.070	0.037	-0.002, 0.142	1.915	.056
	Age not controlled for	0.077	0.085	-0.088, 0.243	0.916	.360
Measurement type	Intercept (directly)	0.072	0.064	-0.053, 0.197	1.133	.257
	Hand scans	0.053	0.106	-0.155, 0.261	0.502	.616

	Self-reported: *s* = 1					
	Unknown: *s* = 0					
Finger injuries	Intercept (finger injuries controlled for)	0.136	0.062	0.014, 0.257	2.190	.029
	Finger injuries not controlled for	-0.104	0.081	-0.262, -0.054	-1.292	.196
Left vs right hand ratios	Intercept (right 2D:4D)	0.070	0.044	-0.017, 0.156	1.584	.113
	Left 2D:4D	0.004	0.004	-0.005, 0.012	0.840	.401
Left vs right hand ratios: Fertility	Intercept (right 2D:4D)	0.031	0.050	-0.067, 0.128	0.618	.537
	Left 2D:4D	0.004	0.004	-0.005, 0.012	0.853	.394

Other moderators with too few *k*/*s*:

Low fertility sample: students vs non-students; Peer-reviewed vs not peer-reviewed; Converted effect size; Non-relevant controls; Number of measurements; Left vs right hand.

Note. *k* = number of observations, MAT = mating, REP = reproductive, *s* = number of samples. Moderation analyses were only run where each level of the moderator included observations from at least two studies and three independent samples. Analyses were run on the mating measures mating behaviours and mating attitudes, and the reproductive measures fertility and reproductive success when there were enough observations to do so.

Appendix E.4

Voice pitch: moderation analyses. The intercept shows the 'simple effect' for the reference category (specified) and the moderator effect shows the change in effect size for that category relative to the reference category. Moderators are bolded if significant after controlling for multiple comparisons, as indicated by computation of q-values. The full list of q-values can be found in Supplementary File 7.

Mating vs reproductive domain

		B	SE	[95% CI]	*z*	*p*
Domain type	Intercept (mating domain)	0.132	0.037	0.061, 0.204	3.610	<.001
	Reproductive domain	0.004	0.075	-0.143, 0.151	0.059	.953

Mating domain (MAT), mating behaviours & mating attitudes

Moderators with too few *k*/*s* (i.e., all potential moderators):

MAT measure type: behaviours vs attitudes; Sample type: low vs high fertility; Low fertility sample: predominantly students vs non-students; High fertility sample: traditional vs industrialised; Ethnicity: predominantly white vs not; Marriage system: monogamy vs non-monogamy; Publication status: published results; Peer-reviewed; Sexual orientation: heterosexual sample vs gay/mixed/unknown; Normality-transformed variables; Converted effect size; Age controlled for; Non-relevant controls; Sex of experimenter; Illness; Smoker; Condition: baseline vs competition vs courtship.

Reproductive domain (REP), fertility & reproductive success

Moderators with too few *k*/*s* (i.e., all potential moderators):

REP measure type: reproductive success vs fertility; Sample type: low vs high fertility; Low fertility sample: predominantly students vs non-students; High fertility sample: traditional vs industrialised; Ethnicity: predominantly white vs not; Marriage system: monogamy vs non-monogamy; Publication status: published results; Peer-reviewed; Sexual orientation: heterosexual sample vs gay/mixed/unknown; Normality-transformed variables; Converted effect size; Age controlled for; Non-relevant controls; Sex of experimenter; Illness; Smoker; Condition: baseline vs competition vs courtship.

Note. *k* = number of observations, MAT = mating, REP = reproductive, *s* = number of samples. Moderation analyses were only run where each level of the moderator included observations from at least two studies and three independent samples. Analyses were run on the mating measures mating behaviours and mating attitudes, and the reproductive measures fertility and reproductive success when there were enough observations to do so.

Appendix E.5

Height: moderation analyses. The intercept shows the 'simple effect' for the reference category (specified) and the moderator effect shows the change in effect size for that category relative to the reference category. Moderators are bolded if significant after controlling for multiple comparisons, as indicated by computation of q-values. The full list of q-values can be found in Supplementary File 7.

Mating vs reproductive domain

		B	SE	[95% CI]	*z*	*p*
Domain type	Intercept (mating domain)	0.049	0.020	0.010, 0.088	2.475	.013
	Reproductive domain	-0.032	0.017	-0.066, 0.001	-1.900	.057

Mating domain (MAT), mating behaviours & mating attitudes

		B	SE	[95% CI]	*z*	*p*
MAT measure type	Intercept (MAT behaviours)	0.054	0.015	0.024, 0.084	3.504	.001
	MAT attitudes	0.000	0.021	-0.041, 0.041	0.007	.995
Sample type	Intercept (low fertility)	0.055	0.016	0.024, 0.086	3.456	.001
	High fertility	0.031	0.062	-0.092, 0.153	0.492	.623
Sample type: MAT behaviours	Intercept (low fertility)	0.051	0.018	0.017, 0.086	2.923	.004
	High fertility	0.034	0.063	-0.090, 0.158	0.541	.588
Low fertility sample	Intercept (predominantly students)	0.065	0.024	0.018, 0.111	2.722	.007
	Non-students/mixed/unknown	-0.018	0.033	-0.082, 0.047	-0.537	.591
Low fertility sample: MAT attitudes	Intercept (predominantly students)	0.045	0.053	-0.060, 0.150	0.843	.399
	Non-students/mixed/unknown	-0.020	0.058	-0.135, 0.094	-0.346	.729
Low fertility sample: MAT behaviours	Intercept (predominantly students)	0.057	0.026	0.006, 0.109	2.171	.030
	Non-students/mixed/unknown	-0.010	0.037	-0.082, 0.062	-0.284	.777
Ethnicity	Intercept (predominantly white)	0.073	0.023	0.028, 0.118	3.153	.002
	Mixed/other/unknown	-0.029	0.031	-0.090, 0.032	-0.918	.359
Ethnicity: MAT behaviours	Intercept (predominantly white)	0.072	0.026	0.021, 0.123	2.761	.006
	Mixed/other/unknown	-0.031	0.035	-0.098, 0.037	-0.890	.374
Publication status	Intercept (published results)	0.067	0.030	0.009, 0.126	2.269	.023

	Non-published results	-0.014	0.035	-0.082, 0.054	-0.403	.687
Publication status: MAT behaviours	Intercept (published results)	0.074	0.030	0.016, 0.133	2.486	.013
	Non-published results	-0.030	0.036	-0.100, 0.040	-0.833	.405
Peer-review	Intercept (peer-reviewed)	0.055	0.016	0.023, 0.087	3.348	.001
	Not peer-reviewed	0.031	0.058	-0.083, 0.145	0.534	.593
Peer-review: MAT behaviours	Intercept (peer-reviewed)	0.050	0.018	0.016, 0.085	2.867	.004
	Not peer-reviewed	0.052	0.066	-0.076, 0.181	0.799	.424
Sexual orientation	Intercept (heterosexual sample)	0.041	0.021	-0.000, 0.082	1.954	.051
	Gay/mixed/unknown	0.038	0.032	-0.025, 0.101	1.180	.238
Sexual orientation: MAT behaviours	Intercept (heterosexual sample)	0.035	0.023	-0.011, 0.081	1.507	.132
	Gay/mixed/unknown	0.042	0.035	-0.027, 0.111	1.199	.231
Normality-transformed variables	Intercept (non-transformed variables)	0.049	0.020	0.011, 0.087	2.521	.012
	Transformed variables	0.031	0.031	-0.030, 0.091	0.995	.320
Normality-transformed variables: MAT behaviours	Intercept (non-transformed variables)	0.039	0.022	-0.004, 0.081	1.796	.073
	Transformed variables	0.053	0.036	-0.017, 0.122	1.480	.139
Converted effect size	Intercept (not converted)	0.052	0.017	0.019, 0.086	3.068	.002
	Converted	0.048	0.043	-0.036, 0.133	1.118	.264
Converted effect size: MAT behaviours	Intercept (not converted)	0.046	0.019	0.010, 0.082	2.481	.013
	Converted	0.055	0.044	-0.031, 0.141	1.244	.214
Age control	Intercept (age controlled for)	0.056	0.018	0.021, 0.090	3.160	.002
	Age not controlled for	0.013	0.018	-0.022, 0.048	0.708	.479
Age control: MAT behaviours	Intercept (age controlled for)	0.051	0.017	0.018, 0.084	2.993	.003

	Age not controlled for	0.012	0.018	-0.023, 0.047	0.663	.507
Number of measurements	Intercept (unknown number of measurements)	0.047	0.021	0.006, 0.088	2.255	.024
	2 measurements	0.096	0.057	-0.015, 0.207	1.702	.089
	1 measurement: *s* = 1					
Number of measurements: MAT behaviours	Intercept (unknown number of measurements)	0.051	0.023	0.006, 0.095	2.231	.026
	2 measurements	0.088	0.070	-0.049, 0.226	1.264	.206
	1 measurement: *s* = 1					
Measurement type	Intercept (measured)	0.057	0.021	0.016, 0.098	2.705	.007
	Self-reported	0.001	0.036	-0.069, 0.071	0.035	.972
	Unknown measurement type	0.010	0.067	-0.122, 0.142	0.144	.886
Measurement type: MAT behaviours	Intercept (measured)	0.054	0.022	0.011, 0.097	2.455	.014
	Self-reported	-0.003	0.040	-0.080, 0.075	-0.073	.942
	Unknown measurement type	0.026	0.079	-0.128, 0.180	0.329	.742

Other moderators with too few *k*/*s*:

High fertility sample: traditional vs industrialised; Marriage system: monogamy vs non-monogamy; Non-relevant controls.

Reproductive domain (REP), fertility & reproductive success

		B	SE	[95% CI]	*z*	*p*
REP measure type	Intercept (reproductive success)	-0.031	0.053	-0.135, 0.073	-0.584	.559
	Fertility	0.045	0.054	-0.061, 0.150	0.831	.406
Sample type	Intercept (low fertility)	-0.037	0.044	-0.123, 0.050	-0.825	.409
	High fertility	0.071	0.057	-0.041, 0.182	1.243	.214
Sample type: Fertility	Intercept (low fertility)	-0.037	0.035	-0.105, 0.031	-1.060	.289
	High fertility	0.090	0.048	-0.004, 0.185	1.878	.060
High fertility sample	Intercept (traditional)	0.029	0.051	-0.071, 0.130	0.575	.565
	Industrialised	0.011	0.081	-0.147, 0.170	0.141	.888
High fertility sample: Fertility	Intercept (traditional)	0.048	0.042	-0.035, 0.131	1.132	.258

	Industrialised	0.016	0.057	-0.096, 0.129	0.287	.774
Ethnicity	Intercept (predominantly white)	0.038	0.047	-0.054, 0.130	0.808	.419
	Mixed/other/unknown	-0.049	0.059	-0.165, 0.066	-0.839	.401
Ethnicity: Fertility	Intercept (predominantly white)	0.037	0.040	-0.042, 0.116	0.924	.355
	Mixed/other/unknown	-0.044	0.053	-0.147, 0.059	-0.838	.402
Marriage system	Intercept (monogamy)	-0.010	0.034	-0.078, 0.057	-0.304	.761
	Non-monogamy	0.052	0.060	-0.066, 0.169	0.862	.389
Marriage system: Fertility	Intercept (monogamy)	-0.001	0.030	-0.059, 0.057	-0.034	.973
	Non-monogamy	0.052	0.061	-0.067, 0.171	0.860	.390
Publication status	Intercept (published results)	-0.023	0.038	-0.098, 0.052	-0.601	.548
	Non-published results	0.064	0.056	-0.046, 0.175	1.139	.255
Publication status: Fertility	Intercept (published results)	-0.018	0.036	-0.088, 0.052	-0.501	.617
	Non-published results	0.063	0.052	-0.040, 0.165	1.202	.229
Sexual orientation	Intercept (heterosexual sample)	-0.092	0.048	-0.185, 0.001	-1.939	.053
	Gay/mixed/unknown	**0.135**	**0.056**	**0.026, 0.245**	**2.417**	**.016**
Sexual orientation: Fertility	Intercept (heterosexual sample)	-0.070	0.041	-0.151, 0.011	-1.702	.089
	Gay/mixed/unknown	0.117	0.050	0.019, 0.214	2.342	.019
Normality-transformed variables	Intercept (non-transformed variables)	-0.005	0.033	-0.069, 0.060	-0.149	.881
	Transformed variables	0.049	0.065	-0.078, 0.175	0.756	.450
Normality-transformed variables: Fertility	Intercept (non-transformed variables)	-0.002	0.030	-0.060, 0.056	-0.081	.936
	Transformed variables	0.067	0.064	-0.059, 0.193	1.036	.300
Converted effect size	Intercept (not converted)	0.022	0.034	-0.045, 0.089	0.645	.519
	Converted	-0.048	0.058	-0.161, 0.065	-0.833	.405
Converted effect size: Fertility	Intercept (not converted)	0.033	0.032	-0.029, 0.095	1.051	.293

	Converted	-0.072	0.058	-0.186, 0.041	-1.246	.213
Converted effect size: Reproductive success	Intercept (not converted)	-0.005	0.132	-0.264, 0.254	-0.037	.971
	Converted	-0.086	0.187	-0.452, 0.280	-0.463	.644
Age control	Intercept (age controlled for)	0.005	0.034	-0.061, 0.071	0.154	.878
	Age not controlled for	-0.032	0.083	-0.195, 0.131	-0.386	.699
Non-relevant controls	Intercept (no non-relevant controls)	0.012	0.032	-0.050, 0.075	0.387	.699
	Non-relevant controls	-0.029	0.068	-0.161, 0.103	-0.428	.668
Non-relevant controls: Fertility	Intercept (no non-relevant controls)	0.017	0.030	-0.041, 0.076	0.585	.559
	Non-relevant controls	-0.033	0.073	-0.176, 0.111	-0.443	.657
Non-relevant controls: Reproductive success	Intercept (no non-relevant controls)	0.017	0.116	-0.210, 0.244	0.150	.881
	Non-relevant controls	-0.166	0.184	-0.526, 0.194	-0.902	.367
Measurement type	Intercept (measured)	-0.007	0.039	-0.083, 0.070	-0.167	.867
	Self-reported	0.012	0.065	-0.116, 0.140	0.178	.859
	Unknown: *s* = 2					
Measurement type: Fertility	Intercept (measured)	-0.001	0.038	-0.076, 0.073	-0.037	.971
	Self-reported	0.006	0.058	-0.108, 0.121	0.105	.916
	Unknown: *s* = 2					

Other moderators with too few *k*/*s*:

Low fertility sample: predominantly students vs non-students; Peer-reviewed; Number of measurements.

Note. *k* = number of observations, MAT = mating, REP = reproductive, *s* = number of samples. Moderation analyses were only run where each level of the moderator included observations from at least two studies and three independent samples. Analyses were run on the mating measures mating behaviours and mating attitudes, and the reproductive measures fertility and reproductive success when there were enough observations to do so.

Appendix E.6

Testosterone levels: moderation analyses. The intercept shows the 'simple effect' for the reference category (specified) and the moderator effect shows the change in effect size for that category relative to the reference category. Moderators are bolded if significant after controlling for multiple comparisons, as indicated by computation of q-values. The full list of q-values can be found in Supplementary File 7.

Mating vs reproductive domain

		B	SE	[95% CI]	*z*	*p*
Domain type	Intercept (Mating domain)	0.093	0.014	0.067, 0.120	6.893	<.001
	Reproductive domain	-0.063	0.060	-0.181, 0.054	-1.059	.290

Mating domain (MAT), mating behaviours & mating attitudes

		B	SE	[95% CI]	*z*	*p*
MAT measure type	Intercept (MAT behaviours)	0.087	0.016	0.056, 0.118	5.460	<.001
	MAT attitudes	0.015	0.027	-0.038, 0.068	0.569	.569
Low fertility sample	Intercept (predominantly students)	0.096	0.020	0.057, 0.136	4.755	<.001
	Non-students/mixed/unknown	0.012	0.041	-0.069, 0.094	0.298	.766
Low fertility sample: MAT behaviours	Intercept (predominantly students)	0.080	0.024	0.032, 0.128	3.288	.001
	Non-students/mixed/unknown	0.024	0.039	-0.053, 0.101	0.610	.542
Ethnicity	Intercept (predominantly white)	0.104	0.023	0.058, 0.150	4.436	<.001
	Mixed/other/unknown	-0.016	0.030	-0.074, 0.043	-0.530	.596
Ethnicity: MAT attitudes	Intercept (predominantly white)	0.126	0.060	0.008, 0.245	2.090	.037
	Mixed/other/unknown	-0.045	0.074	-0.189, 0.099	-0.614	.540
Ethnicity: MAT behaviours	Intercept (predominantly white)	0.099	0.022	0.056, 0.142	4.497	<.001
	Mixed/other/unknown	-0.024	0.025	-0.074, 0.026	-0.943	.346
Publication status	Intercept (published results)	0.097	0.017	0.064, 0.130	5.736	<.001

	Non-published results	0.012	0.039	-0.088, 0.065	0.303	.762
Publication status: MAT behaviours	Intercept (published results)	0.082	0.014	0.054, 0.110	5.703	<.001
	Non-published results	0.019	0.041	-0.061, 0.098	0.465	.642
Sexual orientation	Intercept (heterosexual sample)	0.125	0.016	0.092, 0.157	7.578	<.001
	Gay/mixed/unknown	**-0.059**	**0.020**	**-0.098, -0.021**	**-2.994**	**.003**
Sexual orientation: MAT attitudes	Intercept (heterosexual sample)	0.165	0.059	0.050, 0.281	2.803	.005
	Gay/mixed/unknown	-0.108	0.076	-0.256, 0.041	-1.419	.156
Sexual orientation: MAT behaviours	Intercept (heterosexual sample)	0.110	0.020	0.071, 0.149	5.529	<.001
	Gay/mixed/unknown	-0.042	0.024	-0.089, 0.005	-1.751	.080
Normality-transformed variables	Intercept (non-transformed variables)	0.073	0.010	0.053, 0.093	7.172	<.001
	Transformed variables	**0.057**	**0.023**	**0.011, 0.103**	**2.445**	**.015**
Normality-transformed variables: MAT behaviours	Intercept (non-transformed variables)	0.074	0.012	0.050, 0.098	5.995	<.001
	Transformed variables	0.036	0.028	-0.018, 0.091	1.306	.192
Converted effect size	Intercept (not converted)	0.091	0.016	0.061, 0.121	5.863	<.001
	Converted	0.029	0.043	-0.056, 0.114	0.665	.506
Converted effect size: MAT behaviours	Intercept (not converted)	0.078	0.013	0.054, 0.103	6.227	<.001
	Converted	0.058	0.046	-0.032, 0.147	1.266	.206
Age control	Intercept (age controlled for)	0.098	0.022	0.054, 0.142	4.393	<.001
	Age not controlled for	-0.011	0.032	-0.074, 0.052	-0.346	.730

Age control: MAT behaviours	Intercept (age controlled for)	0.107	0.029	0.050, 0.163	3.690	<.001
	Age not controlled for	-0.029	0.041	-0.109, 0.051	-0.708	.479
Non-relevant controls	Intercept (no non-relevant controls)	0.088	0.015	0.059, 0.117	5.954	<.001
	Non-relevant controls	0.042	0.041	-0.039, 0.122	1.012	.312
Time of day	Intercept (AM)	0.090	0.023	0.045, 0.135	3.899	<.001
	PM	0.016	0.028	-0.038, 0.070	0.568	.570
	Mixed/unknown	-0.014	0.046	-0.105, 0.076	-0.309	.757
Time of day: MAT attitudes	Intercept (AM)	0.069	0.045	-0.019, 0.157	1.543	.123
	PM	0.038	0.047	-0.055, 0.130	0.794	.427
	Mixed/unknown: *s* = 2					
Time of day: MAT behaviours	Intercept (AM)	0.088	0.024	0.041, 0.134	3.672	<.001
	PM	0.003	0.036	-0.068, 0.074	0.087	.931
	Mixed/unknown	-0.017	0.049	-0.114, 0.079	-0.350	.726
Blood contamination	Intercept (checked)	0.142	0.031	0.080, 0.203	4.533	<.001
	Not checked/unknown	-0.061	0.036	-0.133, 0.010	-1.694	.090
Blood contamination: MAT behaviours	Intercept (checked)	0.130	0.032	0.067, 0.193	4.036	<.001
	Not checked/unknown	-0.064	0.040	-0.141, 0.014	-1.599	.110
Fatherhood status	Intercept (non-fathers)	0.069	0.025	0.021, 0.117	2.831	.005
	Mixed/unknown	0.040	0.031	-0.021, 0.101	1.286	.198
	Fathers: *s* = 0					
Fatherhood status: MAT attitudes	Intercept (non-fathers)	0.072	0.054	-0.033, 0.177	1.339	.181
	Mixed/unknown	0.061	0.081	-0.097, 0.219	0.755	.450
	Fathers: *s* = 0					

Fatherhood status: MAT behaviours	Intercept (non-fathers)	0.055	0.031	-0.005, 0.115	1.792	.073
	Mixed/unknown	0.043	0.036	-0.027, 0.114	1.203	.229
	Fathers: *s* = 0					
Other moderators with too few *k*/*s*:						
Sample type: low vs high fertility; High fertility sample: traditional vs industrialised; Marriage system: monogamy vs non-monogamy; Peer-reviewed; How assayed: blood vs saliva; Relationship status						
Reproductive domain (REP), fertility & reproductive success						
Moderators with too few *k*/*s* (i.e., all potential moderators):						
REP measure type (reproductive success vs fertility; Sample type: low vs high fertility; Low fertility sample: predominantly students vs non-students; High fertility sample: traditional vs industrialised; Ethnicity: predominantly white vs not; Marriage system: monogamy vs non-monogamy; Publication status: published results; Peer-reviewed; Sexual orientation: heterosexual sample vs gay/mixed/unknown; Normality-transformed variables; Converted effect size; Age controlled for; Non-relevant controls; How assayed: blood vs saliva; Time of day: AM vs PM vs mixed/unknown; Fatherhood status; Relationship status						

Note. *k* = number of observations, MAT = mating, REP = reproductive, *s* = number of samples. Moderation analyses were only run where each level of the moderator included observations from at least two studies and three independent samples. Analyses were run on the mating measures mating behaviours and mating attitudes, and the reproductive measures fertility and reproductive success when there were enough observations to do so.

Appendix F.1

Mating domain, reproductive domain, and offspring mortality domain predicted by global masculinity. Pearson's r (95% CI); p value for meta-analytic effect, q-value (correcting for multiple comparisons); number of observations (k), samples (s), and unique participants (n); test for heterogeneity (Q), p value for heterogeneity. Statistically significant meta-analytic associations are bolded if still significant after controlling for multiple comparisons.

Sample	Mating domain	Reproductive domain	Offspring mortality domain
All samples	***r* = .090 (0.071, 0.110), *p* < .001, *q* = .001** *k* = 371, *s* = 70, *n* = 117481 Q(df = 370) = 1108.213, *p* < .001	*r* = .047 (0.004, 0.090), *p* = .033, *q* = .080 *k* = 81, *s* = 36, *n* = 107848 Q(df = 80) = 628.883, *p* < .001	*r* = .002 (-0.011, 0.015), *p* = .782, *q* = .475 *k* = 22, *s* = 13, *n* = 21991 Q(df = 21) = 14.765, *p* = .835

Note. k = number of observations; n = number of unique participants; Q = Cochran's Q test of heterogeneity; s = number of samples.

Appendix F.2

Global masculinity: moderation analyses. The intercept shows the 'simple effect' for the reference category (specified) and the moderator effect shows the change in effect size for that category relative to the reference category. Moderators are bolded if significant after controlling for multiple comparisons, as indicated by computation of q-values. The full list of q-values can be found in Supplementary File 7.

Global masculinity						
		B	SE	[95% CI]	*z*	*p*
Domain type	**Intercept (mating domain)**	**0.083**	**0.010**	**0.062, 0.103**	**8.022**	**<.001**
	Reproductive domain	0.004	0.003	-0.003, 0.011	1.209	.227
	Offspring mortality domain	-0.054	0.030	-0.113, 0.005	-1.794	.073

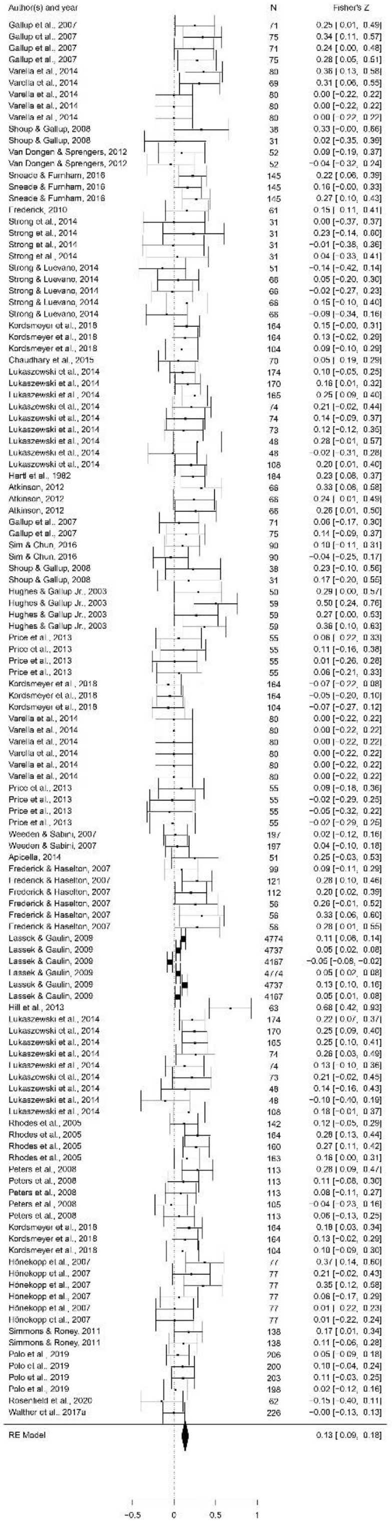
Mating domain: moderation analyses of type of masculinity						
		B	SE	[95% CI]	*z*	*p*
Masculinity type	**Intercept (body masculinity)**	**0.108**	**0.012**	**0.085, 0.132**	**8.951**	**<.001**
	2D:4D	**-0.060**	**0.018**	**-0.096, -0.025**	**-3.325**	**.001**
	Facial masculinity	**-0.054**	**0.023**	**-0.099, -0.009**	**-2.354**	**.019**
	Height	**-0.020**	**0.009**	**-0.037, -0.002**	**-2.235**	**.025**
	Testosterone levels	-0.009	0.020	-0.047, 0.030	-0.452	.652
	Voice pitch	0.033	0.041	-0.047, 0.113	0.814	.416

Reproductive domain: moderation analyses of type of masculinity						
		B	SE	[95% CI]	*z*	*p*
Masculinity type	**Intercept (body masculinity)**	**0.117**	**0.039**	**0.040, 0.194**	**2.980**	**.003**
	2D:4D	-0.042	0.052	-0.143, 0.059	-0.812	.417
	Facial masculinity	-0.030	0.071	-0.170, 0.110	-0.417	.677
	Height	**-0.107**	**0.038**	**-0.181, -0.033**	**-2.817**	**.005**
	Testosterone levels	-0.028	0.069	-0.163, 0.107	-0.403	.687
	Voice pitch	-0.032	0.079	-0.187, 0.122	-0.411	.681

Note. Moderation analyses were only run where each level of the moderator included observations from at least two studies and three independent samples.

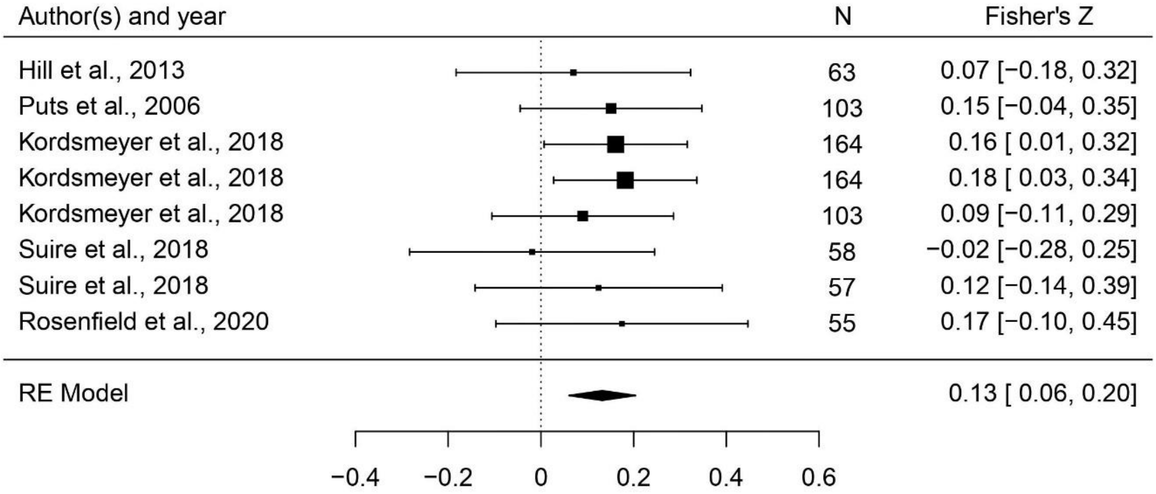
Appendix G.1

Forest plot of the association between body masculinity and the mating domain. Effect sizes are shown as Z-transformed r, with 95% confidence intervals in brackets. The width of the diamond corresponds to the confidence interval for the overall effect.



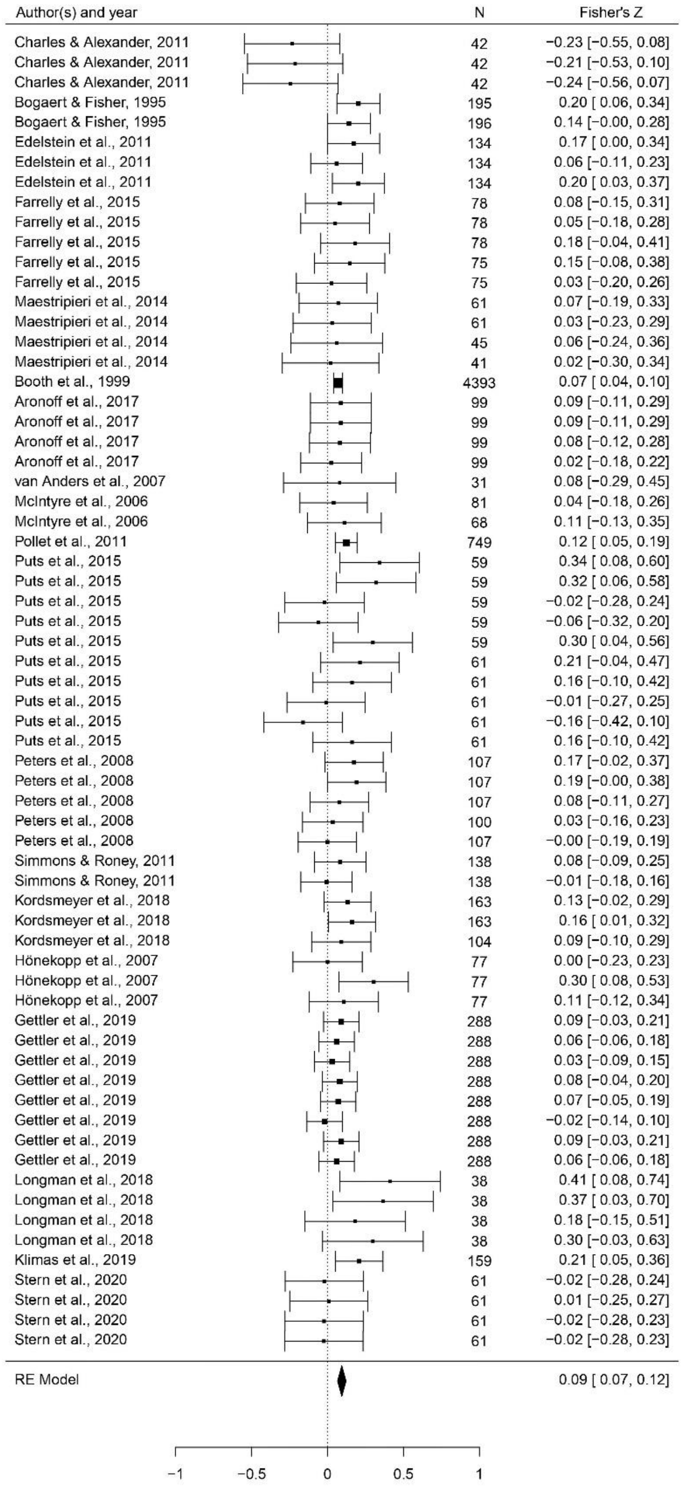
Appendix G.2

Forest plot of the association between voice pitch and the mating domain. Effect sizes are shown as Z-transformed r, with 95% confidence intervals in brackets. The width of the diamond corresponds to the confidence interval for the overall effect.



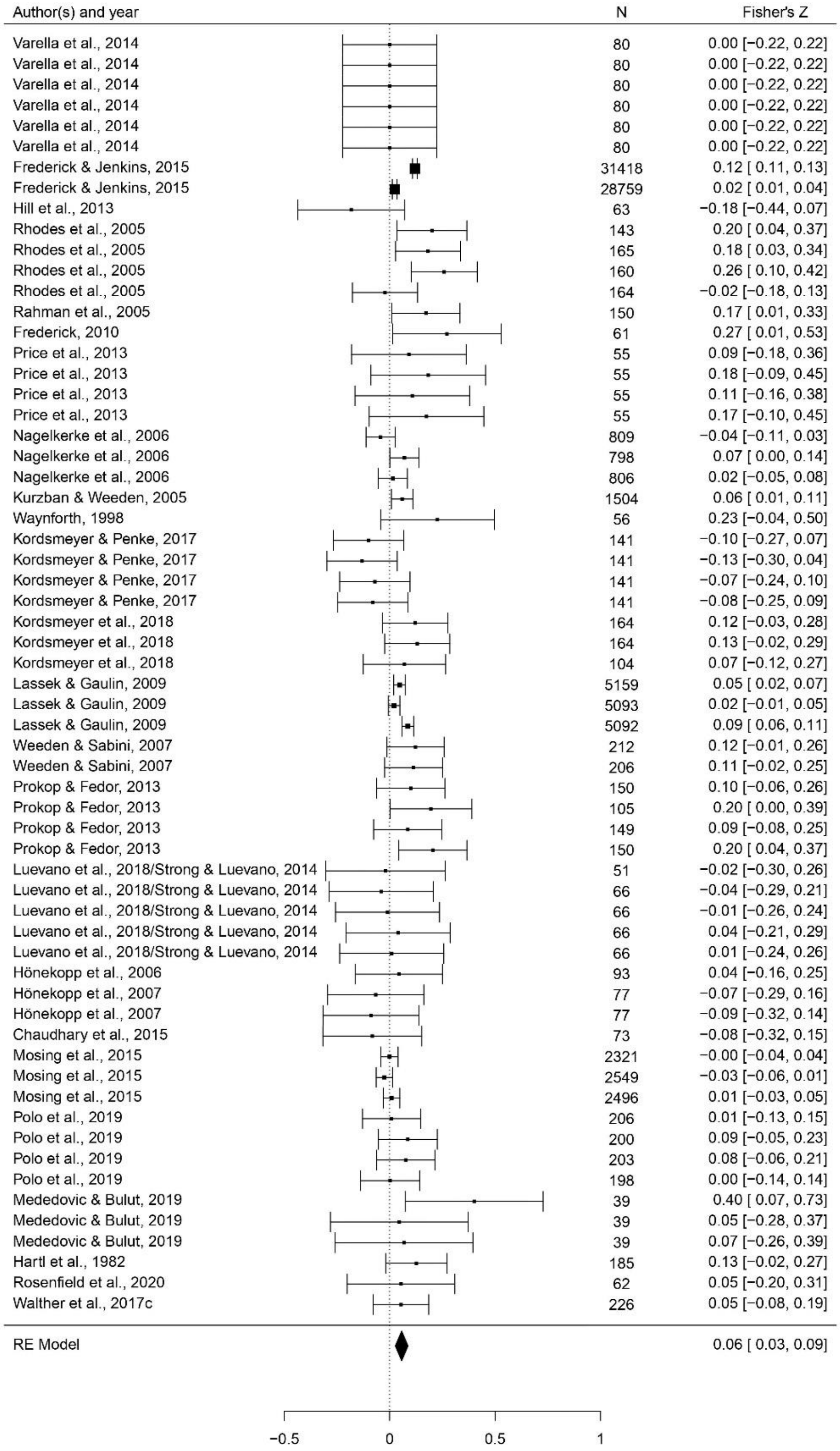
Appendix G.3

Forest plot of the association between testosterone levels and the mating domain. Effect sizes are shown as Z-transformed r, with 95% confidence intervals in brackets. The width of the diamond corresponds to the confidence interval for the overall effect.



Appendix G.4

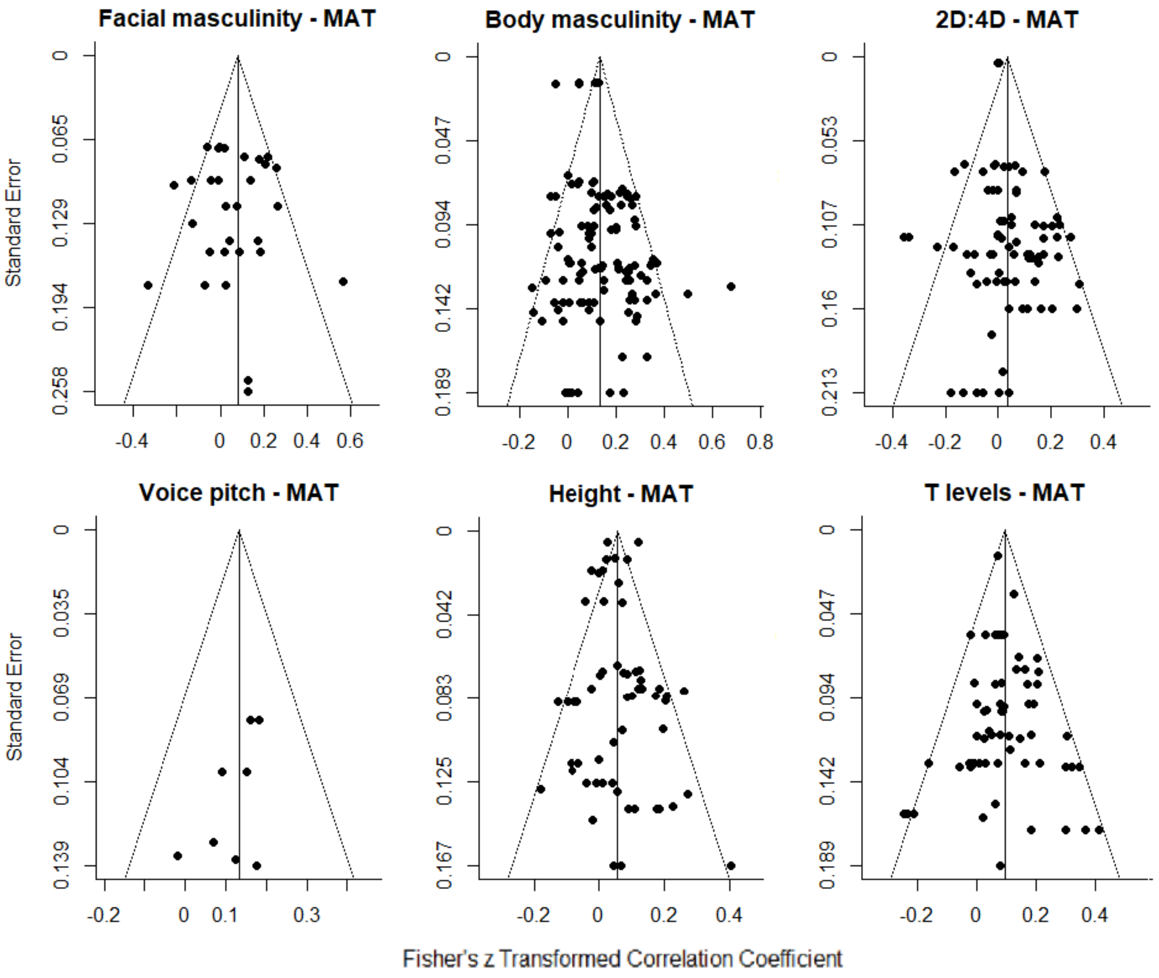
Forest plot of the association between height and the mating domain. Effect sizes are shown as Z-transformed r, with 95% confidence intervals in brackets. The width of the diamond corresponds to the confidence interval for the overall effect.



Author(s) and year	N	Fisher's Z
Varella et al., 2014	80	0.00 [−0.22, 0.22]
Varella et al., 2014	80	0.00 [−0.22, 0.22]
Varella et al., 2014	80	0.00 [−0.22, 0.22]
Varella et al., 2014	80	0.00 [−0.22, 0.22]
Varella et al., 2014	80	0.00 [−0.22, 0.22]
Varella et al., 2014	80	0.00 [−0.22, 0.22]
Frederick & Jenkins, 2015	31418	0.12 [0.11, 0.13]
Frederick & Jenkins, 2015	28759	0.02 [0.01, 0.04]
Hill et al., 2013	63	−0.18 [−0.44, 0.07]
Rhodes et al., 2005	143	0.20 [0.04, 0.37]
Rhodes et al., 2005	165	0.18 [0.03, 0.34]
Rhodes et al., 2005	160	0.26 [0.10, 0.42]
Rhodes et al., 2005	164	−0.02 [−0.18, 0.13]
Rahman et al., 2005	150	0.17 [0.01, 0.33]
Frederick, 2010	61	0.27 [0.01, 0.53]
Price et al., 2013	55	0.09 [−0.18, 0.36]
Price et al., 2013	55	0.18 [−0.09, 0.45]
Price et al., 2013	55	0.11 [−0.16, 0.38]
Price et al., 2013	55	0.17 [−0.10, 0.45]
Nagelkerke et al., 2006	809	−0.04 [−0.11, 0.03]
Nagelkerke et al., 2006	798	0.07 [0.00, 0.14]
Nagelkerke et al., 2006	806	0.02 [−0.05, 0.08]
Kurzban & Weeden, 2005	1504	0.06 [0.01, 0.11]
Waynforth, 1998	56	0.23 [−0.04, 0.50]
Kordsmeyer & Penke, 2017	141	−0.10 [−0.27, 0.07]
Kordsmeyer & Penke, 2017	141	−0.13 [−0.30, 0.04]
Kordsmeyer & Penke, 2017	141	−0.07 [−0.24, 0.10]
Kordsmeyer & Penke, 2017	141	−0.08 [−0.25, 0.09]
Kordsmeyer et al., 2018	164	0.12 [−0.03, 0.28]
Kordsmeyer et al., 2018	164	0.13 [−0.02, 0.29]
Kordsmeyer et al., 2018	104	0.07 [−0.12, 0.27]
Lassek & Gaulin, 2009	5159	0.05 [0.02, 0.07]
Lassek & Gaulin, 2009	5093	0.02 [−0.01, 0.05]
Lassek & Gaulin, 2009	5092	0.09 [0.06, 0.11]
Weeden & Sabini, 2007	212	0.12 [−0.01, 0.26]
Weeden & Sabini, 2007	206	0.11 [−0.02, 0.25]
Prokop & Fedor, 2013	150	0.10 [−0.06, 0.26]
Prokop & Fedor, 2013	105	0.20 [0.00, 0.39]
Prokop & Fedor, 2013	149	0.09 [−0.08, 0.25]
Prokop & Fedor, 2013	150	0.20 [0.04, 0.37]
Luevano et al., 2018/Strong & Luevano, 2014	51	−0.02 [−0.30, 0.26]
Luevano et al., 2018/Strong & Luevano, 2014	66	−0.04 [−0.29, 0.21]
Luevano et al., 2018/Strong & Luevano, 2014	66	−0.01 [−0.26, 0.24]
Luevano et al., 2018/Strong & Luevano, 2014	66	0.04 [−0.21, 0.29]
Luevano et al., 2018/Strong & Luevano, 2014	66	0.01 [−0.24, 0.26]
Hönekopp et al., 2006	93	0.04 [−0.16, 0.25]
Hönekopp et al., 2007	77	−0.07 [−0.29, 0.16]
Hönekopp et al., 2007	77	−0.09 [−0.32, 0.14]
Chaudhary et al., 2015	73	−0.08 [−0.32, 0.15]
Mosing et al., 2015	2321	−0.00 [−0.04, 0.04]
Mosing et al., 2015	2549	−0.03 [−0.06, 0.01]
Mosing et al., 2015	2496	0.01 [−0.03, 0.05]
Polo et al., 2019	206	0.01 [−0.13, 0.15]
Polo et al., 2019	200	0.09 [−0.05, 0.23]
Polo et al., 2019	203	0.08 [−0.06, 0.21]
Polo et al., 2019	198	0.00 [−0.14, 0.14]
Mededovic & Bulut, 2019	39	0.40 [0.07, 0.73]
Mededovic & Bulut, 2019	39	0.05 [−0.28, 0.37]
Mededovic & Bulut, 2019	39	0.07 [−0.26, 0.39]
Hartl et al., 1982	185	0.13 [−0.02, 0.27]
Rosenfield et al., 2020	62	0.05 [−0.20, 0.31]
Walther et al., 2017c	226	0.05 [−0.08, 0.19]
RE Model		0.06 [0.03, 0.09]

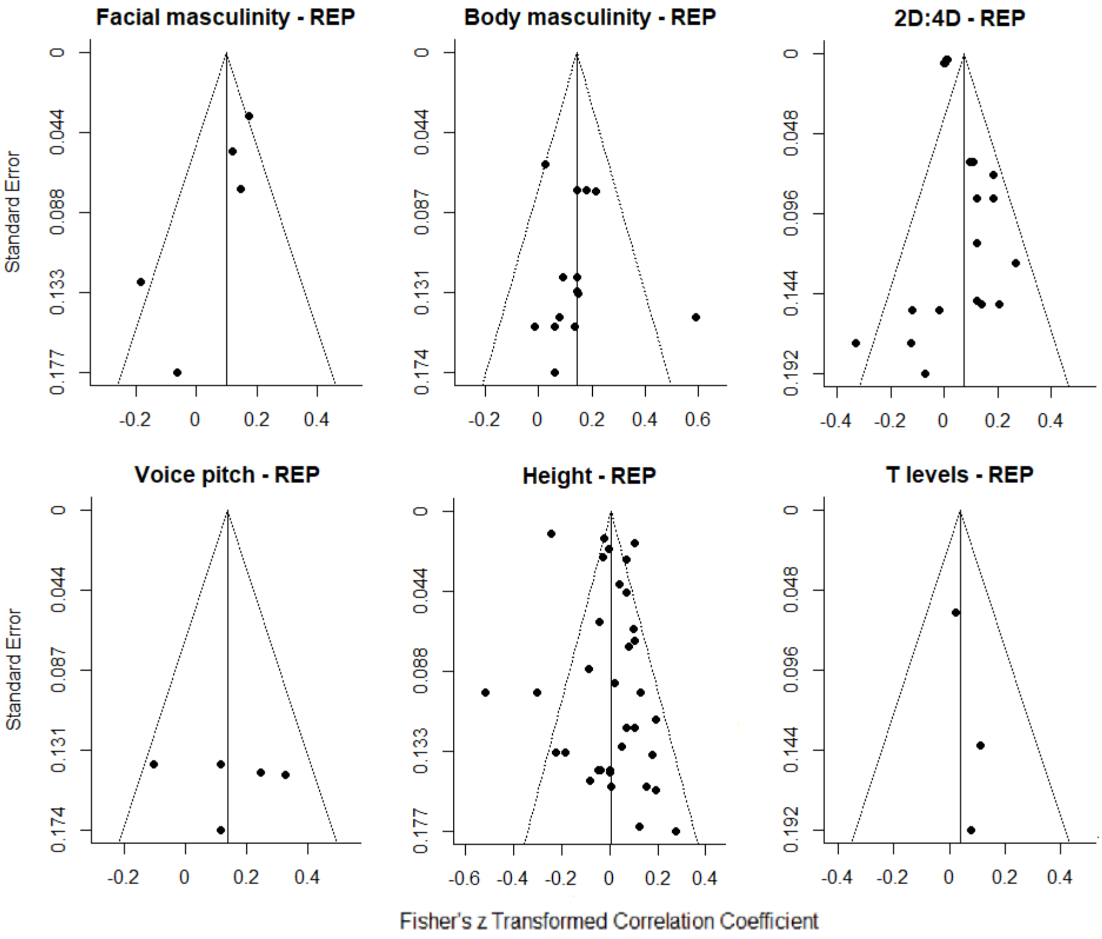
Appendix H.1

Funnel plots of effect sizes for mating measures (MAT). T = testosterone.



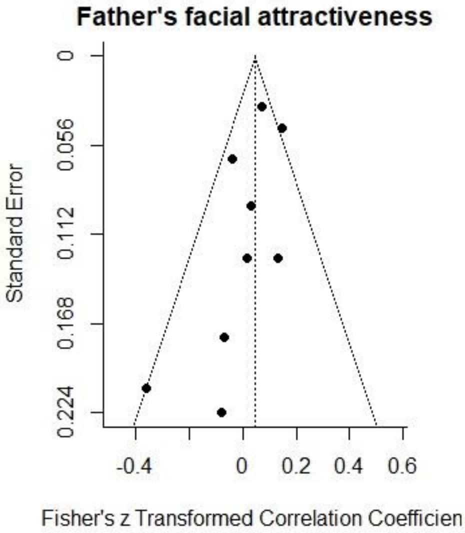
Appendix H.2

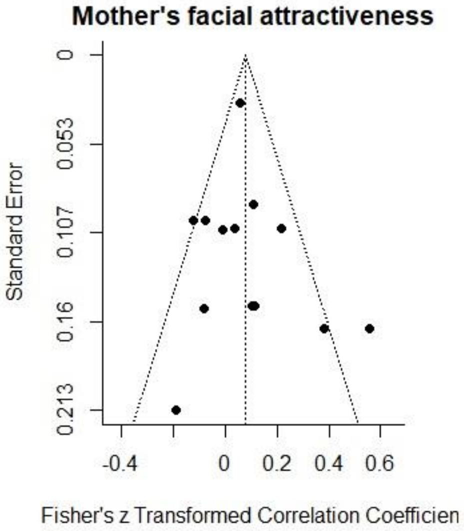
Funnel plots of effect sizes for reproductive measures (REP). T = testosterone levels.



Appendix I

Funnel plots of effect sizes for parents' facial attractiveness and reproductive measures





Appendix J

Offspring health questions (Study 4)

1. How many colds/bouts of flu has your child suffered in the past three years?
2. How many stomach upsets (gastroenteritis/food poisoning /diarrhoea and vomiting bugs) has your child suffered in the past three years?
3. On how many occasions has your child required antibiotics in the three years?
4. Does your child suffer from any long running conditions? (please check all that apply)

   Allergies (other)

   Asthma

   Coeliac disease

   Diabetes

   Eczema

   Hay fever

   Psoriasis
5. During the LAST THREE MONTHS on how many days has your child experienced the following:

   Mild cold

   Severe cold/flu

   Stomach infection/vomiting

   Taken antibiotics

Appendix K

9 items adapted from The Adolescent Sexual Activity Index (Hansen et al., 1999)

Have you engaged in any of the following activities with a romantic partner (e.g., a boy/girlfriend or someone you've 'pulled')?

(1) hugging yes/no

(2) holding hands yes/no

(3) spending time alone together yes/no

(4) kissing yes/no

(5) cuddling yes/no

(6) laying down together yes/no

(7) having someone put his or her hands under your clothing yes/no

(8) putting your hands under someone else's clothing yes/no

(9) being naked together yes/no

Appendix L

Hollingshead Four-Factor Index of Socioeconomic Status (SES; Hollingshead, 1979), adapted for use in the U.K.

Parent's education code is rated on a 7-point scale that lists highest grade completed, in which 7 = graduate/professional training, 6 = standard college or university graduation, 5 = partial college, at least one year of specialised training, 4 = high school graduate, 3 = partial high school, 10th or 11th grade, 2 = junior high school, including 9th grade, 1 = less than 7th grade, 0 = not applicable or unknown. The child participant's parent's occupational code is rated on a 9-point scale, for which the Hollingshead manuscript provides a more detailed list: 9 = higher executive, proprietor of large businesses, major professional, 8 = administrators, lesser professionals, proprietor of medium-sized business, 7 = smaller business owners, farm owners, managers, minor professionals, 6 = technicians, semi-professionals, small business owners (business valued at $50,000-70,000), 5 = clerical and sales workers, small farm and business owners (business valued at $25,000-50,000), 4 = smaller business owners (<$25,000), skilled manual laborers, craftsmen, tenant farmers, 3 = machine operators and semi-skilled workers, 2 = unskilled workers, 1 = farm laborers, menial service workers, students, housewives, (dependent on welfare, no regular occupation), 0 = not applicable or unknown. An SES score is then calculated for a total parental SES score.

Appendix M

Pubertal Development Scale (PDS; Petersen et al., 1988)

This questionnaire is about your physical development and asks about your height, whether you have started getting spots, whether your body hair has begun to grow, etc. Please answer the following questions honestly. If you do not feel comfortable answering a question, please leave it blank.

1. Would you say that your growth in height has not yet begun to spurt, has barely started, is definitely underway, or does growth seem completed? (spurt = more growth than usual) No/Yes (Barely)/Yes (Definitely)/Development completed
2. And how about the growth of your body hair? Would you say that your body hair has not yet started growing, has barely started growing, or does growth seem completed? (body hair=underarm or pubic hair)

   No/Yes (Barely)/Yes (Definitely)/Development completed
3. Have you noticed any skin changes, especially spots on your face?

   No/Yes (Barely)/Yes (Definitely)
4. IF YOU ARE A BOY: Have you noticed a deepening of your voice?

   No/Yes (Barely)/Yes (Definitely)/Development completed
5. IF YOU ARE A BOY: Have you begun to grow hair on your face?

   No/Yes (Barely)/Yes (Definitely)/Development completed
6. IF YOU ARE GIRL: Have your breasts begun to grow?

   No/Yes (Barely)/Yes (Definitely)/Development completed
7. IF YOU ARE A GIRL: Have you begun to menstruate (get your period)?

   No/Yes (Barely)/Yes (Definitely)

   7b. How old were you when you got your period? (this question was not used)

Appendix N

Holmes and Rahe Stress Scale for Non-Adults (Holmes & Rahe, 1967)

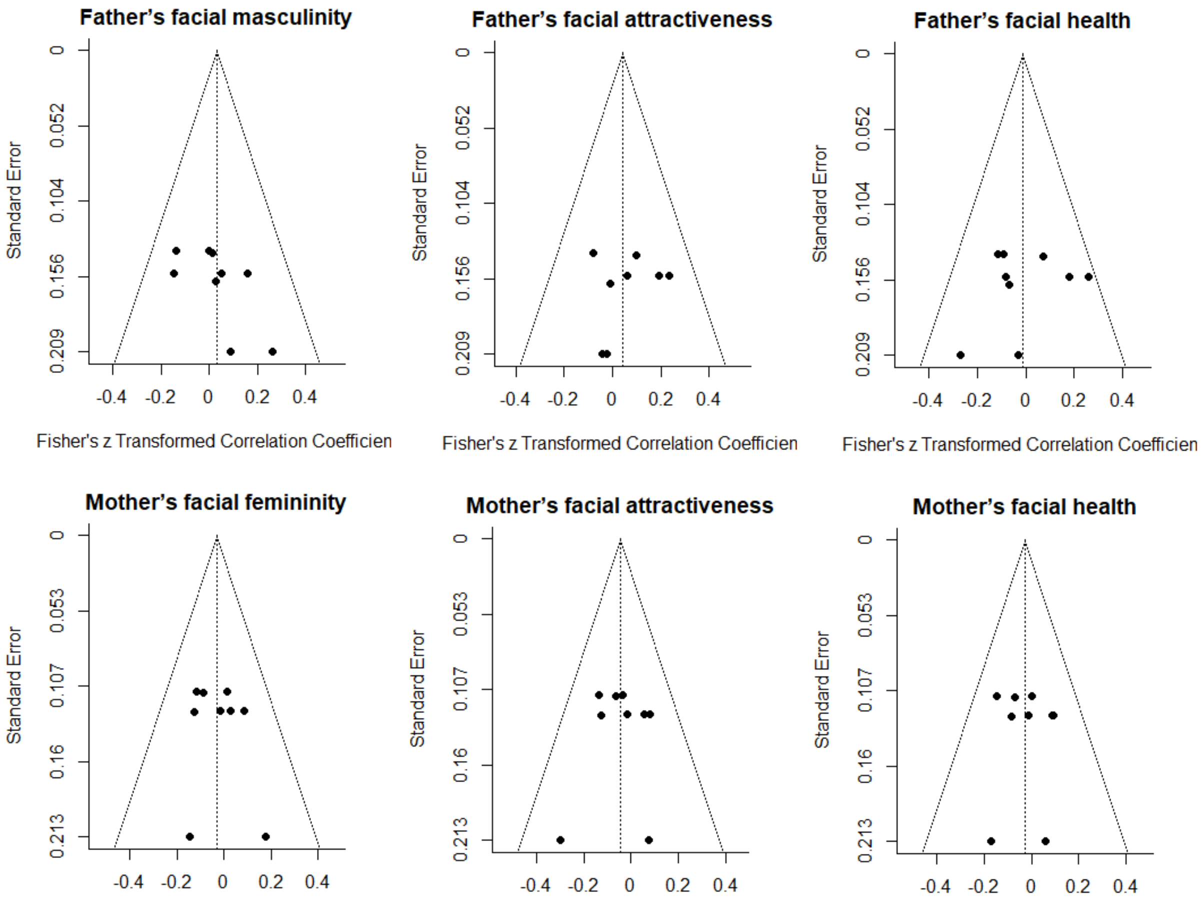
How many of the following have you/your child experienced in the last five (in child dataset) or two (in teenager dataset) years?

Death of parent	100
Unplanned pregnancy/abortion	100
Getting married	95
Divorce of parents	90
Acquiring a visible deformity	80
Fathering a child	70
Jail sentence of parent for over one year	70
Marital separation of parents	69
Death of a brother or sister	68
Change in acceptance by peers	67
Unplanned pregnancy of sister	64
Discovery of being an adopted child	63
Marriage of parent to stepparent	63
Death of a close friend	63
Having a visible congenital deformity	62
Serious illness requiring hospitalization	58
Failure of a grade in school	56
Not making an extracurricular activity	55
Hospitalisation of a parent	55
Jail sentence of parent for over 30 days	53

Breaking up with boyfriend or girlfriend	53
Beginning to date	51
Suspension from school	50
Becoming involved with drugs or alcohol	50
Birth of a brother or sister	50
Increase in arguments between parents	47
Loss of job by parent	46
Outstanding personal achievement	46
Change in parent's financial status	45
Accepted at college of choice	43
Being a senior in high school	42
Hospitalisation of a sibling	41
Increased absence of parent from home	38
Brother or sister leaving home	37
Addition of third adult to family	34
Becoming a full fledged member of a church	31
Decrease in arguments between parents	27
Decrease in arguments with parents	26
Mother or father beginning work	26

Appendix O

Funnel plots of effect sizes for parental traits and offspring health


Father's facial masculinity

Father's facial attractiveness

Father's facial health

Mother's facial femininity

Mother's facial attractiveness

Mother's facial health

Appendix P

Self-Perceived Mating Success Scale (Landolt et al., 1995)

1. Members of the opposite sex that I like, tend to like me back
2. Members of the opposite sex notice me
3. I receive many compliments from members of the opposite sex
4. Members of the opposite sex are not very attracted to me (reverse coded)
5. I receive sexual invitations from members of the opposite sex
6. Members of the opposite sex are attracted to me
7. I can have as many sexual partners as I choose
8. I do not receive many compliments from members of the opposite sex (reverse coded)

Appendix Q

Between-sex Pearson's bivariate correlations between traits in the student couples, controlling (where relevant) for men's facial hair and women's makeup use (Study 7). Male traits are shown in the rows and female traits in the columns.

	Female facial fem.	Female facial attr.	Female facial health	Female facial dom.	Female facial prestige	Female HGS	Female height	Female BMI	Female WHR
Male facial masc.	.059 *N* =97	.104 *N* = 97	.088 *N* = 97	.151 *N* = 97	.158 *N* = 97	.127 *N* = 99	-.039 *N* = 99	.137 *N* = 93	-.001 *N* = 99
Male facial attr.	.021 *N* = 97	.084 *N* = 97	.195 *N* = 97	-.134 *N* = 97	.085 *N* = 97	-.024 *N* = 99	-.132 *N* = 99	-.069 *N* = 93	-.073 *N* = 99
Male facial health	.053 *N* = 97	.005 *N* = 97	.140 *N* = 97	-.214* *N* = 97	-.021 *N* = 97	-.104 *N* = 99	-.239* *N* = 99	-.153 *N* = 93	-.056 *N* = 99
Male facial dom.	.072 *N* = 97	.122 *N* = 97	.159 *N* = 97	.104 *N* = 97	.161 *N* = 97	.064 *N* = 99	-.105 *N* = 99	.066 *N* = 93	-.052 *N* = 99
Male facial prestige	-.031 *N* = 97	-.052 *N* = 97	.070 *N* = 97	-.065 *N* = 97	-.052 *N* = 97	-.173 *N* = 99	-.113 *N* = 99	.020 *N* = 93	.030 *N* = 99
Male HGS	.134 *N* = 98	.227* *N* = 98	.217* *N* = 98	.036 *N* = 98	.226* *N* = 98	.229* *N* = 100	.207* *N* = 100	-.007 *N* = 94	-.124 *N* = 100
Male height	.156 *N* = 98	.179 *N* = 98	.191 *N* = 98	.127 *N* = 98	.212* *N* = 98	.093 *N* = 100	.205* *N* = 100	.063 *N* = 94	-.026 *N* = 100
Male BMI	.218* *N* = 91	.184 *N* = 91	-.004 *N* = 91	-.009 *N* = 91	.155 *N* = 91	-.058 *N* = 93	-.077 *N* = 93	.252* *N* = 88	.114 *N* = 93
Male shoulder circ.	.215* *N* = 97	.205* *N* = 97	.059 *N* = 97	.044 *N* = 97	.157 *N* = 97	.000 *N* = 99	.038 *N* = 99	.165 *N* = 93	.053 *N* = 99

Note. Attr = attractiveness; circ = circumference; dom = dominance; fem = femininity; HGS = handgrip strength; masc = masculinity; WHR = waist-to-hip ratio.
* $p < .05$. No associations remained significant after computation of q-values.

Appendix R

Between-sex Pearson's bivariate correlations between traits in the British parent couples, controlling (where relevant) for age, men's facial hair and women's makeup use (Study 8). Male traits are shown in the rows and female traits in the columns.

	Female facial fem.	Female facial attr.	Female facial health	Female facial dom.	Female HGS	Female height
Male facial masc.	.147 *N* = 58	-.012 *N* = 58	-.048 *N* = 58	.065 *N* = 58	.062 *N* = 44	-.237 *N* = 43
Male facial attr.	**.359**** ***N* = 58**	**.430**** ***N* = 58**	**.367**** ***N* = 58**	-.113 *N* = 58	.233 *N* = 44	.130 *N* = 43
Male facial health	**.446***** ***N* = 58**	**.478***** ***N* = 58**	**.413**** ***N* = 58**	-.161 *N* = 58	.181 *N* = 44	.125 *N* = 43
Male facial dom.	.074 *N* = 58	.025 *N* = 58	-.026 *N* = 58	.149 *N* = 58	-.117 *N* = 44	-.363* *N* = 43
Male HGS	.254 *N* = 31	.226 *N* = 31	.259 *N* = 31	-.235 *N* = 31	-.069 *N* = 29	.269 *N* = 29
Male height	.131 *N* = 31	.074 *N* = 31	.070 *N* = 31	-.161 *N* = 31	-.149 *N* = 29	.428* *N* = 29

Note. Attr = attractiveness; dom = dominance; fem = femininity; HGS = handgrip strength; masc = masculinity.
*** *p* < .001, ** *p* < .01, * *p* < .05. Associations that remained significant after computation of *q*-values are bolded.

References

Acar, H., & Eler, N. (2018). The relationship of digit ratio (2D:4D) with cerebral lateralization and grip strength in elite swimmers. *Journal of Education and Training Studies, 6*(4), 84-89.

Albert, G., Arnocky, S., Puts, D. A., & Hodges-Simeon, C. R. (2021a). Can listeners assess men’s self-reported health from their voice?. *Evolution and Human Behavior, 42*(2), 91-103.

Albert, G., Wells, E., Arnocky, S., Liu, C. H., & Hodges-Simeon, C. R. (2021b). Observers use facial masculinity to make physical dominance assessments following 100-ms exposure. *Aggressive Behavior, 47*(2), 226-235.

Alharbi, S. A., 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, 6*(4), 328-334.

Allen, H. L., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F., ... & Ferreira, T. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature, 467*(7317), 832.

Alvergne, A., Faurie, C., & Raymond, M. (2009). Variation in testosterone levels and male reproductive effort: Insight from a polygynous human population. *Hormones and Behavior, 56*(5), 491-497.

Anderson, C., Hildreth, J. A. D., & Howland, L. (2015). Is the desire for status a fundamental human motive? A review of the empirical literature. *Psychological Bulletin*, *141*(3), 574-601.

Andersson, M. B. (1994). *Sexual selection.* Princeton University Press.

Apicella, C. L. (2014). Upper-body strength predicts hunting reputation and reproductive success in Hadza hunter-gatherers. *Evolution and Human Behavior, 35*(6), 508-518.

Apicella, C. L., Feinberg, D. R., & Marlowe, F. W. (2007). Voice pitch predicts reproductive success in male hunter-gatherers. *Biology Letters, 3*(6), 682-684.

Apicella, C. L., Tobolsky, V. A., Marlowe, F. W., & Miller, K. W. (2016). Hadza hunter-gatherer men do not have more masculine digit ratios (2D:4D). *American Journal of Physical Anthropology, 159*(2), 223-232.

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

Archer, J. (2019). The reality and evolutionary significance of human psychological sex differences. *Biological Reviews*, *94*(4), 1381-1415.

Arnocky, S., Carré, J. M., Bird, B. M., Moreau, B. J., Vaillancourt, T., Ortiz, T., & Marley, N. (2018a). The facial width-to-height ratio predicts sex drive, sociosexuality, and intended infidelity. *Archives of Sexual Behavior, 47*(5), 1375-1385.

Arnocky, S., Hodges-Simeon, C. R., Ouellette, D., & Albert, G. (2018b). Do men with more masculine voices have better immunocompetence?. *Evolution and Human Behavior, 39*(6), 602-610.

Aronoff, J. (2017). *Life history strategies, testosterone, and the anthropology of human development.* [Master’s thesis, The University of Alabama]. ProQuest Dissertations and Theses database.

Arslanian, S., & Suprasongsin, C. (1997). Testosterone treatment in adolescents with delayed puberty: Changes in body composition, protein, fat, and glucose metabolism. *The Journal of Clinical Endocrinology & Metabolism, 82*(10), 3213-3220.

Atherton, P. J., & Smith, K. (2012). Muscle protein synthesis in response to nutrition and exercise. *The Journal of Physiology, 590*(5), 1049-1057.

Atkinson, J. A. (2012). *Anthropometric correlates of reproductive success, facial configuration, risk taking and sexual behaviours among indigenous and Western populations: The role of hand-grip strength and wrist width.* [Doctoral thesis, University at Albany].

Atkinson, J., Pipitone, R. N., Sorokowska, A., Sorokowski, P., Mberira, M., Bartels, A., & Gallup Jr, G. G. (2012). Voice and handgrip strength predict reproductive success in a group of indigenous African females. *PLoS One, 7*(8), e41811.

Aung, T., & Puts, D. (2020). Voice pitch: a window into the communication of social power. *Current Opinion in Psychology, 33*, 154-161.

Bailey, D. H., Durante, K. M., & Geary, D. C. (2011). Men’s perception of women's attractiveness is calibrated to relative mate value and dominance of the women's partner. *Evolution and Human Behavior, 32*(2), 138-146.

Baird, D. D., Wilcox, A. J., Weinberg, C. R., Kamel, F., McConnaughey, D. R., Musey, P. I., & Collins, D. C. (1997). Preimplantation hormonal differences between the conception and non-conception menstrual cycles of 32 normal women. *Human Reproduction (Oxford, England), 12*(12), 2607-2613.

Barber, N. (1995). The evolutionary psychology of physical attractiveness: Sexual selection and human morphology. *Ethology and Sociobiology, 16*(5), 395-424.

Barnard, C. J. (2004). *Animal behaviour: Mechanism, development, function and evolution.* Pearson Education.

Bhasin, S., Storer, T. W., Berman, N., Callegari, C., Clevenger, B., Phillips, J., ... & Casaburi, R. (1996). The effects of supraphysiologic doses of testosterone on muscle size and strength in normal men. *New England Journal of Medicine, 335*(1), 1-7.

Bhasin, S., Woodhouse, L., Casaburi, R., Singh, A. B., Bhasin, D., Berman, N., ... & Storer, T. W. (2001). Testosterone dose-response relationships in healthy young men. *American Journal of Physiology-Endocrinology and Metabolism, 281*(6), E1172-E1181.

Björntorp, P. (1987). The associations between obesity, adipose tissue distribution and disease. *Acta Medica Scandinavica, 222*(S723), 121-134.

Blaker, N. M., & van Vugt, M. (2014). The status-size hypothesis: How cues of physical size and social status influence each other. In *The psychology of social status* (pp. 119-137). Springer, New York, NY.

Blethen, S. L., Gaines, S., & Weldon, V. (1984). Comparison of predicted and adult heights in short boys: Effect of androgen therapy. *Pediatric Research, 18*(5), 467-469.

Boesch, C., Kohou, G., Néné, H., & Vigilant, L. (2006). Male competition and paternity in wild chimpanzees of the Taï forest. *American Journal of Physical Anthropology: The*

Official Publication of the American Association of Physical Anthropologists, 130(1), 103-115.

Bogaert, A. F., & Fisher, W. A. (1995). Predictors of university men's number of sexual partners. *Journal of Sex Research, 32*(2), 119-130.

Booth, A., & Dabbs Jr, J. M. (1993). Testosterone and men's marriages. *Social Forces, 72*(2), 463-477.

Booth, A., Johnson, D. R., & Granger, D. A. (1999). Testosterone and men's health. *Journal of Behavioral Medicine, 22*(1), 1-19.

Boothroyd, L. G., Cross, C. P., Gray, A. W., Coombes, C., & Gregson-Curtis, K. (2011). Perceiving the facial correlates of sociosexuality: Further evidence. *Personality and Individual Differences, 50*(3), 422-425.

Boothroyd, L. G., Gray, A. W., Headland, T. N., Uehara, R. T., Waynforth, D., Burt, D. M., & Pound, N. (2017). Male facial appearance and offspring mortality in two traditional societies. *PloS One, 12*(1), e0169181.

Boothroyd, L. G., Jones, B. C., Burt, D. M., Cornwell, R. E., Little, A. C., Tiddeman, B. P., & Perrett, D. I. (2005). Facial masculinity is related to perceived age but not perceived health. *Evolution and Human Behavior, 26*(5), 417-431.

Boothroyd, L. G., Jones, B. C., Burt, D. M., DeBruine, L. M., & Perrett, D. I. (2008). Facial correlates of sociosexuality. *Evolution and Human Behavior, 29*(3), 211-218.

Boothroyd, L. G., Jones, B. C., Burt, D. M., & Perrett, D. I. (2007). Partner characteristics associated with masculinity, health and maturity in male faces. *Personality and Individual Differences, 43*(5), 1161-1173.

Boothroyd, L. G., Jucker, J. L., Thornborrow, T., Tovée, M. J., Batres, C., & Penton-Voak, I. (2021). Testing mate choice hypotheses in a transitional small scale population. *Adaptive Human Behavior and Physiology, 7*(3), 220-244.

Boothroyd, L. G., Scott, I., Gray, A. W., Coombes, C. I., & Pound, N. (2013). Male facial masculinity as a cue to health outcomes. *Evolutionary Psychology, 11*(5), 147470491301100508.

Bordini, B., & Rosenfield, R. L. (2011). Normal pubertal development. *Pediatrics in Review*, *32*(6), 223-229.

Borgerhoff Mulder, M. (2009). Serial monogamy as polygyny or polyandry?: Marriage in the Tanzanian Pimbwe. *Human Nature: An Interdisciplinary Biosocial Perspective, 20*(2), 130-150.

Botwin, M. D., Buss, D. M., & Shackelford, T. K. (1997). Personality and mate preferences: Five factors in mate selection and marital satisfaction. *Journal of Personality, 65*(1), 107-136.

Bovet, J. (2019). Evolutionary theories and men's preferences for women's waist-to-hip ratio: Which hypotheses remain? A systematic review. *Frontiers in Psychology*, *10*, 1221.

Boyd, R., & Richerson, P. J. (1982). Cultural transmission and the evolution of cooperative behavior. *Human Ecology*, *10*(3), 325-351.

Boyette, A. H., Lew-Levy, S., Sarma, M. S., & Gettler, L. T. (2019). Testosterone, fathers as providers and caregivers, and child health: Evidence from fisher-farmers in the Republic of the Congo. *Hormones and Behavior, 107*, 35-45.

Bribiescas, R. G. (1996). Testosterone levels among Aché hunter-gatherer men. *Human Nature*, *7*(2), 163-188.

Brooks, R. C., Shelly, J. P., Jordan, L. A., & Dixson, B. J. (2015). The multivariate evolution of female body shape in an artificial digital ecosystem. *Evolution and Human Behavior*, *36*(5), 351-358.

Brown, G. R., Laland, K. N., & Mulder, M. B. (2009). Bateman's principles and human sex roles. *Trends in Ecology & Evolution, 24*(6), 297-304.

Bulygina, E., Mitteroecker, P., & Aiello, L. (2006). Ontogeny of facial dimorphism and patterns of individual development within one human population. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists, 131*(3), 432-443.

Burch, R. L. (2020). More than just a pretty face: The overlooked contributions of women in evolutionary psychology textbooks. *Evolutionary Behavioral Sciences, 14*(1), 100-114.

Burriss, R. P., Roberts, S. C., Welling, L. L., Puts, D. A., & Little, A. C. (2011). Heterosexual romantic couples mate assortatively for facial symmetry, but not masculinity. *Personality and Social Psychology Bulletin, 37*(5), 601-613.

Butovskaya, M., Sorokowska, A., Karwowski, M., Sabiniewicz, A., Fedenok, J., Dronova, D., ... & Sorokowski, P. (2017). Waist-to-hip ratio, body-mass index, age and number of children in seven traditional societies. *Scientific Reports*, *7*(1), 1-9.

Butovskaya, M. L., Windhager, S., Karelin, D., Mezentseva, A., Schaefer, K., & Fink, B. (2018). Associations of physical strength with facial shape in an African pastoralist society, the Maasai of Northern Tanzania. *PloS One, 13*(5), e0197738.

Butterfield, S. A., Lehnhard, R. A., Loovis, E. M., Coladarci, T., Saucier, D. (2009). Grip strength performances by 5- to 19-year-olds. *Perceptual and Motor Skills, 109*(2), 362-370.

Cai, Z., Hahn, A. C., Zhang, W., Holzleitner, I. J., Lee, A. J., DeBruine, L. M., & Jones, B. C. (2019). No evidence that facial attractiveness, femininity, averageness, or coloration are cues to susceptibility to infectious illnesses in a university sample of young adult women. *Evolution and Human Behavior, 40*(2), 156-159.

Cartei, V., Bond, R., & Reby, D. (2014). What makes a voice masculine: Physiological and acoustical correlates of women's ratings of men’s vocal masculinity. *Hormones and Behavior, 66*(4), 569-576.

Chagnon, N. A. (1988). Life histories, blood revenge, and warfare in a tribal population. *Science, 239*(4843), 985-992.

Charles, N. E., & Alexander, G. M. (2011). The association between 2D: 4D ratios and sociosexuality: A failure to replicate. *Archives of Sexual Behavior, 40*(3), 587-595.

Chaudhary, N., Salali, G. D., Thompson, J., Dyble, M., Page, A., Smith, D., ... & Migliano, A. B. (2015). Polygyny without wealth: popularity in gift games predicts polygyny in BaYaka Pygmies. *Royal Society Open Science, 2*(5), 150054.

Cheverud, J. M., Dow, M. M., & Leutenegger, W. (1985). The quantitative assessment of phylogenetic constraints in comparative analyses: Sexual dimorphism in body weight among primates. *Evolution, 39*(6), 1335-1351.

Cieri, R. L., Churchill, S. E., Franciscus, R. G., Tan, J., & Hare, B. (2014). Craniofacial feminization, social tolerance, and the origins of behavioral modernity. *Current Anthropology, 55*(4), 419-443.

Clark, A. P. (2004). Self-perceived attractiveness and masculinization predict women's sociosexuality. *Evolution and Human Behavior, 25*(2), 113-124.

Clutton-Brock, T. H., & Harvey, P. H. (1977). Primate ecology and social organization. *Journal of Zoology, 183*(1), 1-39.

Coetzee, V., Barrett, L., Greeff, J. M., Henzi, S. P., Perrett, D. I., & Wadee, A. A. (2007). Common HLA alleles associated with health, but not with facial attractiveness. *PLoS One, 2*(7), e640.

Cornwell, R. E., & Perrett, D. I. (2008). Sexy sons and sexy daughters: the influence of parents' facial characteristics on offspring. *Animal Behaviour, 76*(6), 1843-1853.

Courtiol, A., Picq, S., Godelle, B., Raymond, M., & Ferdy, J. B. (2010). From preferred to actual mate characteristics: The case of human body shape. *PloS One, 5*(9), e13010.

Cowlishaw, G., & Dunbar, R. I. (1991). Dominance rank and mating success in male primates. *Animal Behaviour, 41*(6), 1045-1056.

Crook, J. H. (2017). Sexual selection, dimorphism, and social organization in the primates. In *Sexual selection and the descent of man* (pp. 231-281). Routledge.

Cunningham, M. R. (1986). Measuring the physical in physical attractiveness: Quasi-experiments on the sociobiology of female facial beauty. *Journal of Personality and Social Psychology*, *50*(5), 925-935.

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.

Dabbs Jr, J. M., & Mallinger, A. (1999). High testosterone levels predict low voice pitch among men. *Personality and Individual Differences, 27*(4), 801-804.

Dağ, Z. Ö., & Dilbaz, B. (2015). Impact of obesity on infertility in women. *Journal of the Turkish German Gynecological Association, 16*(2), 111-117.

DeBruine, L. M., Jones, B. C., Little, A. C., Boothroyd, L. G., Perrett, D. I., Penton-Voak, I. S., ... & Tiddeman, B. P. (2006). Correlated preferences for facial masculinity and ideal or actual partner's masculinity. *Proceedings of the Royal Society B: Biological Sciences, 273*(1592), 1355-1360.

Devi, M. R., Kumari, J. R., & Srikumari, C. R. (1985). Fertility and mortality differences in relation to maternal body size. *Annals of Human Biology*, *12*(5), 479-484.

Ditzen, B., Palm-Fischbacher, S., Gossweiler, L., Stucky, L., & Ehlert, U. (2017). Effects of stress on women's preference for male facial masculinity and their endocrine correlates. *Psychoneuroendocrinology, 82*, 67-74.

Dixson, A. F. (2009). *Sexual selection and the origins of human mating systems.* Oxford University Press.

Dixson, A., Dixson, B., & Anderson, M. (2005). Sexual selection and the evolution of visually conspicuous sexually dimorphic traits in male monkeys, apes, and human beings. *Annual Review of Sex Research, 16*(1), 1-19.

Dixson, B. J., Sagata, K., Linklater, W. L., & Dixson, A. F. (2010). Male preferences for female waist-to-hip ratio and body mass index in the highlands of Papua New Guinea. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *141*(4), 620-625.

Docherty, C., Lee, A. J., Hahn, A. C., DeBruine, L. M., & Jones, B. C. (2020). Do more attractive women show stronger preferences for male facial masculinity?. *Evolution and Human Behavior, 41*(4), 312-317.

Doll, L. M., Hill, A. K., Rotella, M. A., Cárdenas, R. A., Welling, L. L., Wheatley, J. R., & Puts, D. A. (2014). How well do men’s faces and voices index mate quality and dominance?. *Human Nature, 25*(2), 200-212.

dos Santos, F. K., Maia, J. A., Gomes, T. N. Q., Daca, T., Madeira, A., Katzmarzyk, P. T., & Prista, A. (2014). Secular trends in growth and nutritional status of Mozambican school-aged children and adolescents. *PloS One, 9*(12), e114068.

Dunsworth, H. M. (2020). Expanding the evolutionary explanations for sex differences in the human skeleton. *Evolutionary Anthropology: Issues, News, and Reviews*, *29*(3), 108-116.

Durante, K. M., & Li, N. P. (2009). Oestradiol level and opportunistic mating in women. *Biology Letters*, *5*(2), 179-182.

Durkee, P. K., Goetz, A. T., & Lukaszewski, A. W. (2018). Formidability assessment mechanisms: Examining their speed and automaticity. *Evolution and Human Behavior, 39*(2), 170-178.

Eaton, S., Eaton, S., & Konner, M. J. (1997). Review paleolithic nutrition revisited: A twelve-year retrospective on its nature and implications. *European Journal of Clinical Nutrition*, *51*(4), 207-216.

Edelstein, R. S., Chopik, W. J., & Kean, E. L. (2011). Sociosexuality moderates the association between testosterone and relationship status in men and women. *Hormones and Behavior, 60*(3), 248-255.

Eisenegger, C., Haushofer, J., & Fehr, E. (2011). The role of testosterone in social interaction. *Trends in Cognitive Sciences, 15*(6), 263-271.

Ellis, L., & Hoskin, A. W. (2020). Evidence for androgenic influences on self-rated health. *Journal of Sexual Medicine and Reproductive Health, 3*(9), 1-9.

Emlen, S. T., & Oring, L. W. (1977). Ecology, sexual selection, and the evolution of mating systems. *Science, 197*(4300), 215-223.

Evans, S., Neave, N., Wakelin, D., & Hamilton, C. (2008). The relationship between testosterone and vocal frequencies in human males. *Physiology & Behavior, 93*(4), 783-788.

Falcon, R. G. (2016). Stay, stray or something in-between? A comment on Wlodarski et al. *Biology Letters, 12*(5), 20151069.

Farrelly, D., Owens, R., Elliott, H. R., Walden, H. R., & Wetherell, M. A. (2015). The effects of being in a "new relationship" on levels of testosterone in men. *Evolutionary Psychology, 13*(1), 147470491501300116.

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*(5), 559-564.

Fechner, P. Y. (2003). The biology of puberty: New developments in sex differences. In *Gender differences at puberty,* 17.

Feingold, A. (1988). Matching for attractiveness in romantic partners and same-sex friends: A meta-analysis and theoretical critique. *Psychological Bulletin, 104*(2), 226.

Fiala, V., Třebický, V., Pazhoohi, F., Leongómez, J. D., Tureček, P., Saribay, S. A., ... & Kleisner, K. Facial attractiveness and preference of sexual dimorphism: A comparison across five populations. *Evolutionary Human Sciences,* 1-37.

Fink, B., Neave, N., & Seydel, H. (2007). Male facial appearance signals physical strength to women. *American Journal of Human Biology, 19*(1), 82-87.

Fisher, C. I., Fincher, C. L., Hahn, A. C., Little, A. C., DeBruine, L. M., & Jones, B. C. (2014). Do assortative preferences contribute to assortative mating for adiposity?. *British Journal of Psychology, 105*(4), 474-485.

Fink, B., Thanzami, V., Seydel, H., & Manning, J. T. (2006). Digit ratio and hand-grip strength in German and Mizos men: Cross-cultural evidence for an organizing effect of prenatal testosterone on strength. *American Journal of Human Biology: The Official Journal of the Human Biology Association, 18*(6), 776-782.

Fletcher, G. J., Simpson, J. A., Campbell, L., & Overall, N. C. (2015). Pair-bonding, romantic love, and evolution: The curious case of Homo sapiens. *Perspectives on Psychological Science, 10*(1), 20-36.

Flinn, M. V., & Low, B. S. (1986). Resource distribution, social competition, and mating patterns in human societies. In *Ecological Aspects of Social Evolution*, D. Rubenstein and R. W. Wrangham, eds.

Folstad, I., & Karter, A. J. (1992). Parasites, bright males, and the immunocompetence handicap. *The American Naturalist, 139*(3), 603-622.

Foo, Y. Z., Nakagawa, S., Rhodes, G., & Simmons, L. W. (2017a). The effects of sex hormones on immune function: A meta-analysis. *Biological Reviews, 92*(1), 551-571.

Foo, Y. Z., Simmons, L. W., Perrett, D. I., Holt, P. G., Eastwood, P. R., & Rhodes, G. (2020). Immune function during early adolescence positively predicts adult facial sexual dimorphism in both men and women. *Evolution and Human Behavior, 41*(3), 199-209.

Foo, Y. Z., Simmons, L. W., & Rhodes, G. (2017b). Predictors of facial attractiveness and health in humans. *Scientific Reports*, *7*(1), 1-12.

Foo, Y. Z., Simmons, L. W., & Rhodes, G. (2017c). The relationship between health and mating success in humans. *Royal Society Open Science*, *4*(1), 160603.

Frederick, M. J. (2010). *Effects of early developmental stress on adult physiology and behavior*. [Doctoral dissertation, State University of New York at Albany]. ProQuest Dissertations and Theses database.

Frederick, D. A., & Haselton, M. G. (2007). Why is muscularity sexy? Tests of the fitness indicator hypothesis. *Personality and Social Psychology Bulletin, 33*(8), 1167-1183.

Frederick, D. A., & Jenkins, B. N. (2015). Height and body mass on the mating market: Associations with number of sex partners and extra-pair sex among heterosexual men and women aged 18-65. *Evolutionary Psychology, 13*(3), 1474704915604563.

Frisancho, A. R., Sanchez, J., Pallardel, D., & Yanez, L. (1973). Adaptive significance of small body size under poor socio-economic conditions in southern Peru. *American Journal of Physical Anthropology, 39*(2), 255-261.

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science, 2*(2), 156-168.

Furman, D., Hejblum, B. P., Simon, N., Jojic, V., Dekker, C. L., Thiébaut, R., ... & Davis, M. M. (2014). Systems analysis of sex differences reveals an immunosuppressive role for testosterone in the response to influenza vaccination. *Proceedings of the National Academy of Sciences*, *111*(2), 869-874.

Galis, F., Ten Broek, C. M., Van Dongen, S., & Wijnaendts, L. C. (2010). Sexual dimorphism in the prenatal digit ratio (2D:4D). *Archives of Sexual Behavior, 39*(1), 57-62.

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

Gangestad, S. W., & Simpson, J. A. (2000). The evolution of human mating: Trade-offs and strategic pluralism. *Behavioral and Brain Sciences*, *23*(4), 573-587.

Gangestad, S. W., & Thornhill, R. (2003). Facial masculinity and fluctuating asymmetry. *Evolution and Human Behavior, 24*(4), 231-241.

Gat, A. (2015). Proving communal warfare among hunter-gatherers: The quasi-rousseauan error. *Evolutionary Anthropology: Issues, News, and Reviews, 24*(3), 111-126.

Geniole, S. N., Denson, T. F., Dixson, B. J., 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.

Genovese, J. E. (2008). Physique correlates with reproductive success in an archival sample of delinquent youth. *Evolutionary Psychology, 6*(3), 147470490800600301.

Gettler, L. T., Agustin, S. S., & Kuzawa, C. W. (2010). Testosterone, physical activity, and somatic outcomes among Filipino males. *American Journal of Physical Anthropology, 142*(4), 590-599.

Gettler, L. T., McDade, T. W., Agustin, S. S., Feranil, A. B., & Kuzawa, C. W. (2014). Testosterone, immune function, and life history transitions in Filipino males (Homo sapiens). *International Journal of Primatology*, *35*(3-4), 787-804.

Gettler, L. T., Kuo, P. X., Rosenbaum, S., Avila, J. L., McDade, T. W., & Kuzawa, C. W. (2019). Sociosexuality, testosterone, and life history status: Prospective associations and longitudinal changes among men in Cebu, Philippines. *Evolution and Human Behavior, 40*(2), 249-258.

Gettler, L. T., McDade, T. W., Feranil, A. B., & Kuzawa, C. W. (2011). Longitudinal evidence that fatherhood decreases testosterone in human males. *Proceedings of the National Academy of Sciences, 108*(39), 16194-16199.

Getty, T. (2002). Signaling health versus parasites. *The American Naturalist, 159*(4), 363-371.

Gildersleeve, K., Haselton, M. G., & Fales, M. R. (2014a). Do women's mate preferences change across the ovulatory cycle? A meta-analytic review. *Psychological Bulletin, 140*(5), 1205-1259.

Gildersleeve, K., Haselton, M. G., & Fales, M. R. (2014b). Meta-analyses and p-curves support robust cycle shifts in women's mate preferences: Reply to Wood and Carden (2014) and Harris, Pashler, and Mickes (2014). *Psychological Bulletin, 140*(5), 1272-1280.

Gildner, T. E. (2018). *Life history tradeoffs between testosterone and immune function among Shuar forager-horticulturalists of Amazonian Ecuador*. [Doctoral dissertation, University of Oregon]. ProQuest Dissertations and Theses database.

Glowacki, L., & Wrangham, R. (2015). Warfare and reproductive success in a tribal population. *Proceedings of the National Academy of Sciences, 112*(2), 348-353.

Goetz, A. T. (2010). The evolutionary psychology of violence. *Psicothema, 22*(1), 15-21.

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.

Gould, S. J. (1966). Allometry and size in ontogeny and phylogeny. *Biological Reviews, 41*(4), 587-638.

Gray, A. W., & Boothroyd, L. G. (2012). Female facial appearance and health. *Evolutionary Psychology, 10*(1), 147470491201000108.

Gray, P. B., & Frederick, D. A. (2012). Body image and body type preferences in St. Kitts, Caribbean: A cross-cultural comparison with US samples regarding attitudes towards muscularity, body fat, and breast size. *Evolutionary Psychology, 10*(3), 147470491201000319.

Gray, P. B., Straftis, A. A., Bird, B. M., McHale, T. S., & Zilioli, S. (2020). Human reproductive behavior, life history, and the challenge hypothesis: A 30-year review, retrospective and future directions. *Hormones and Behavior, 123*, 104530.

Gray, J. P., & Wolfe, L. D. (1980). Height and sexual dimorphism of stature among human societies. *American Journal of Physical Anthropology, 53*(3), 441-456.

Grebe, N. M., Sarafin, R. E., Strenth, C. R., & Zilioli, S. (2019). Pair-bonding, fatherhood, and the role of testosterone: A meta-analytic review. *Neuroscience & Biobehavioral Reviews, 98*, 221-233.

Griggs, R. C., Kingston, W., Jozefowicz, R. F., Herr, B. E., Forbes, G., & Halliday, D. (1989). Effect of testosterone on muscle mass and muscle protein synthesis. *Journal of Applied Physiology, 66*(1), 498-503.

Grillot, R. L., Simmons, Z. L., Lukaszewski, A. W., & Roney, J. R. (2014). Hormonal and morphological predictors of women’s body attractiveness. *Evolution and Human Behavior*, *35*(3), 176-183.

Guerra-Silveira, F., & Abad-Franch, F. (2013). Sex bias in infectious disease epidemiology: Patterns and processes. *PloS One, 8*(4), e62390.

Gülcen, B., Pelin, İ. C., & Özener, E. B. (2021). The craniofacial indicators of aggression: A cross-sectional multiparametric anthropometry study. *Folia Morphologica, 80*(1), 55 -62.

Hahn, A. C., Holzleitner, I. J., Lee, A. J., Kandrik, M., O'Shea, K. J., DeBruine, L. M., & Jones, B. C. (2019). Facial masculinity is only weakly correlated with handgrip strength in young adult women. *American Journal of Human Biology, 31*(1), e23203.

Halpern, C. T., Udry, J. R., Campbell, B., & Suchindran, C. (1993). Testosterone and pubertal development as predictors of sexual activity: A panel analysis of adolescent males. *Psychosomatic Medicine, 55*(5), 436.

Hamilton, L. D., Carré, J. M., Mehta, P. H., Olmstead, N., & Whitaker, J. D. (2015). Social neuroendocrinology of status: A review and future directions. *Adaptive Human Behavior and Physiology, 1*(2), 202-230.

Hamilton, W. D., & Zuk, M. (1982). Heritable true fitness and bright birds: A role for parasites?. *Science, 218*(4570), 384-387.

Hammer, M. F., Mendez, F. L., Cox, M. P., Woerner, A. E., & Wall, J. D. (2008). Sex-biased evolutionary forces shape genomic patterns of human diversity. *PLoS Genetics, 4*(9), e1000202.

Hansen, W. B., Paskett, E. D., & Carter, L. J. (1999). The Adolescent Sexual Activity Index (ASAI): A standardized strategy for measuring interpersonal heterosexual behaviours among youth. *Health Education Research, 14*(4), 485-490.

Harries, M., Hawkins, S., Hacking, J., & Hughes, I. (1998). Changes in the male voice at puberty: Vocal fold length and its relationship to the fundamental frequency of the voice. *The Journal of Laryngology & Otology, 112*(5), 451-454.

Harris, C. R., Pashler, H., & Mickes, L. (2014). Elastic analysis procedures: An incurable (but preventable) problem in the fertility effect literature. Comment on Gildersleeve, Haselton, and Fales (2014). *Psychological Bulletin, 140*(5), 1260-1264.

Hartl, E. M., Monnelly, E. P., & Elderkin, R. D. (2013). *Physique and delinquent behavior: A thirty-year follow-up of William H. Sheldon's Varieties of Delinquent Youth*. Academic Press.

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.

Headland, T. N. (1986). *Why foragers do not become farmers: A historical study of a changing ecosystem and its effect on a negrito hunter-gatherer group in the Philippines (agta, cultural ecology, homicide, competitive exclusion principle, time allocation)*. [Doctoral dissertation, University of Hawai'i at Manoa]. ProQuest Dissertations Publishing.

Headland, T. N. (1989). Population decline in a Philippine Negrito hunter-gatherer society. *American Journal of Human Biology, 1*(1), 59-72.

Headland, T., Headland, J., and Uehara, R. (2011). Agta Demographic Database: Chronicle of a hunter-gatherer community in transition. *SIL Language and Culture Documentation and Description, 2.*

Helle, S. (2010). Digit length ratio (2D:4D) and variation in key life-history traits and fitness in contemporary Finnish women. *Behavioral Ecology*, *21*(5), 1061-1066.

Henderson, J. J., & Anglin, J. M. (2003). Facial attractiveness predicts longevity. *Evolution and Human Behavior*, *24*(5), 351-356.

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

Hill, A. K., Bailey, D. H., & Puts, D. A. (2017). Gorillas in our midst? Human sexual dimorphism and contest competition in men. In *On human nature* (pp. 235-249). Academic Press.

Hill, A. K., Hunt, J., Welling, L. L., Cárdenas, R. A., Rotella, M. A., Wheatley, J. R., ... & Puts, D. A. (2013). Quantifying the strength and form of sexual selection on men's traits. *Evolution and Human Behavior, 34*(5), 334-341.

Hill, K., & Hurtado, A. M. (1996). *Demographic/life history of Ache foragers*. Hawthorne, NY.

Hodges-Simeon, C. R., Asif, S., Gurven, M., Blackwell, A. D., & Gaulin, S. J. (2019). Testosterone is positively and estradiol negatively associated with mucosal immunity in Amazonian adolescents. *American Journal of Human Biology, 31*(5), e23284.

Hodges-Simeon, C. R., Gaulin, S. J., & Puts, D. A. (2011). Voice correlates of mating success in men: Examining “contests” versus “mate choice” modes of sexual selection. *Archives of Sexual Behavior, 40*(3), 551-557.

Hodges-Simeon, C. R., Grail, G. P., Albert, G., Landry, N., Ortiz, T. L., Carré, J. M., ... & Arnocky, S. A. (2020). Testosterone, cortisol, and secretory immunoglobulin-A within a single day and across two sequential days among trans-and cis-gender men. *Steroids, 160*, 108640.

Hodges-Simeon, C. R., Hanson Sobraske, K. N., Samore, T., Gurven, M., & Gaulin, S. J. (2016). Facial width-to-height ratio (fWHR) is not associated with adolescent testosterone levels. *PloS One, 11*(4), e0153083.

Hollier, L. P., Keelan, J. A., Jamnadass, E. S., Maybery, M. T., Hickey, M., & Whitehouse, A. J. (2015). Adult digit ratio (2D:4D) is not related to umbilical cord androgen or estrogen concentrations, their ratios or net bioactivity. *Early Human Development, 91*(2), 111-117.

Hollingshead, A. G. (1979). Hollingshead Four-Factor Index of Socio-Economic Status. *New Haven: Yale University, CT.*

Holmboe, S. A., Priskorn, L., Jørgensen, N., Skakkebaek, N. E., Linneberg, A., Juul, A., & Andersson, A. M. (2017). Influence of marital status on testosterone levels–A ten year follow-up of 1113 men. *Psychoneuroendocrinology, 80*, 155-161.

Homish, G. G., & Leonard, K. E. (2008). Spousal influence on general health behaviours in a community sample. *American Journal of Health Behavior, 32*(6), 754-763.

Hume, D. K., & Montgomerie, R. (2001). Facial attractiveness signals different aspects of “quality” in women and men. *Evolution and Human Behavior*, *22*(2), 93-112.

Hoppler, S., Walther, A., La Marca-Ghaemmaghami, P., & Ehlert, U. (2018). Lower birthweight and left-/mixed-handedness are associated with intensified age-related sex steroid decline in men. Findings from the Men's Health 40+ Study. *Andrology, 6*(6), 896-902.

Hughes, S. M., & Gallup Jr, G. G. (2003). Sex differences in morphological predictors of sexual behavior: Shoulder to hip and waist to hip ratios. *Evolution and Human Behavior, 24*(3), 173-178.

Hume, D. K., & Montgomerie, R. (2001). Facial attractiveness signals different aspects of “quality” in women and men. *Evolution and Human Behavior*, *22*(2), 93-112.

Hönekopp, J., Bartholdt, L., Beier, L., & Liebert, A. (2007). Second to fourth digit length ratio (2D:4D) and adult sex hormone levels: New data and a meta-analytic review. *Psychoneuroendocrinology, 32*(4), 313-321.

Hönekopp, J., Rudolph, U., Beier, L., Liebert, A., & Müller, C. (2007). Physical attractiveness of face and body as indicators of physical fitness in men. *Evolution and Human Behavior, 28*(2), 106-111.

Hönekopp, J., Voracek, M., & Manning, J. T. (2006). 2nd to 4th digit ratio (2D:4D) and number of sex partners: Evidence for effects of prenatal testosterone in men. *Psychoneuroendocrinology, 31*(1), 30-37.

Isaac, J. L. (2005). Potential causes and life-history consequences of sexual size dimorphism in mammals. *Mammal Review, 35*(1), 101-115.

Isen, J., McGue, M., & Iacono, W. (2014). Genetic influences on the development of grip strength in adolescence. *American Journal of Physical Anthropology, 154*(2), 189-200.

Jansson, L., & Holmdahl, R. (1998). Estrogen-mediated immunosuppression in autoimmune diseases. *Inflammation Research*, *47*(7), 290-301.

Jasieńska, G., Ziomkiewicz, A., Ellison, P. T., Lipson, S. F., & Thune, I. (2004). Large breasts and narrow waists indicate high reproductive potential in women. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 271*(1545), 1213-1217.

Jennions, M. D., Kahn, A. T., Kelly, C. D., & Kokko, H. (2012). Meta-analysis and sexual selection: Past studies and future possibilities. *Evolutionary Ecology, 26*(5), 1119-1151.

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.

Johnston, V. S., & Franklin, M. (1993). Is beauty in the eye of the beholder?. *Ethology and Sociobiology*, *14*(3), 183-199.

Jokela, M. (2009). Physical attractiveness and reproductive success in humans: evidence from the late 20th century United States. *Evolution and Human Behavior, 30*(5), 342-350.

Jones, B. C., Feinberg, D. R., DeBruine, L. M., Little, A. C., & Vukovic, J. (2010). A domain specific opposite-sex bias in human preferences for manipulated voice pitch. *Animal Behaviour, 79*(1), 57-62.

Jones, B. C., Hahn, A. C., Fisher, C. I., Wang, H., Kandrik, M., Han, C., ... & DeBruine, L. M. (2018). No compelling evidence that preferences for facial masculinity track changes in women’s hormonal status. *Psychological Science*, *29*(6), 996-1005.

Jordão, J. M., & Oliveira, R. F. (2001). Major claws make male fiddler crabs more conspicuous to visual predators: A test using human observers. In *Advances in decapod crustacean research* (pp. 241-247). Springer, Dordrecht.

Jünger, J., Motta-Mena, N. V., Cardenas, R., Bailey, D., Rosenfield, K. A., Schild, C., ... & Puts, D. A. (2018). Do women's preferences for masculine voices shift across the ovulatory cycle?. *Hormones and Behavior, 106,* 122-134.

Kalick, S. M., Zebrowitz, L. A., Langlois, J. H., & Johnson, R. M. (1998). Does human facial attractiveness honestly advertise health? Longitudinal data on an evolutionary question. *Psychological Science, 9*(1), 8-13.

Kallman, D. A., Plato, C. C., & Tobin, J. D. (1990). The role of muscle loss in the age-related decline of grip strength: Cross-sectional and longitudinal perspectives. *Journal of Gerontology, 45*(3), M82-M88.

Kandrik, M., Hahn, A. C., Han, C., Wincenciak, J., Fisher, C. I., DeBruine, L. M., & Jones, B. C. (2017). Does the interaction between cortisol and testosterone predict men's facial attractiveness?. *Adaptive Human Behavior and Physiology, 3*(4), 275-281.

Kaplan, H., Gurven, M., Hill, K., & Hurtado, A. M. (2005). The natural history of human food sharing and cooperation: A review and a new multi-individual approach to the negotiation of norms. *Moral sentiments and material interests: The foundations of cooperation in economic life, 6,* 75-113.

Kasperk, C., Helmboldt, A., Börcsök, I., Heuthe, S., Cloos, O., Niethard, F., & Ziegler, R. (1997). Skeletal site-dependent expression of the androgen receptor in human osteoblastic cell populations. *Calcified Tissue International, 61*(6), 464-473.

Kelly, C. D., Stoehr, A. M., Nunn, C., Smyth, K. N., & Prokop, Z. M. (2018). Sexual dimorphism in immunity across animals: A meta-analysis. *Ecology Letters*, *21*(12), 1885-1894.

Kendall, M. G. (1970). *Rank correlation methods* (4th ed.). Charles Griffin & Co.

Kirchengast, S. (2000). Differential reproductive success and body size in! Kung San people from northern Namibia. Collegium *Antropologicum*, *24*(1), 121-132.

Kirchengast, S., & Winkler, E. M. (1995). Differential reproductive success and body dimensions in Kavango males from urban and rural areas in northern Namibia. *Human Biology, 67*(2), 291-309.

Kleisner, K., Tureček, P., Roberts, S. C., Havlíček, J., Valentova, J. V., Akoko, R. M., ... & Saribay, S. A. (2021). How and why patterns of sexual dimorphism in human faces vary across the world. *Scientific Reports, 11*(1), 1-14.

Klimas, C., Ehlert, U., Lacker, T. J., Waldvogel, P., & Walther, A. (2019). Higher testosterone levels are associated with unfaithful behavior in men. *Biological Psychology*, *146*, 107730.

Klimek, M., Galbarczyk, A., Nenko, I., Alvarado, L. C., & Jasienska, G. (2014). Digit ratio (2D:4D) as an indicator of body size, testosterone concentration and number of children in human males. *Annals of Human Biology, 41*(6), 518-523.

Klimek, M., Galbarczyk, A., Nenko, I., & Jasienska, G. (2016). Women with more feminine digit ratio (2D:4D) have higher reproductive success. *American Journal of Physical Anthropology*, *160*(3), 549-553.

Kordsmeyer, T. L., Freund, D., Pita, S. R., Jünger, J., & Penke, L. (2019). Further evidence that facial width-to-height ratio and global facial masculinity are not positively associated with testosterone levels. *Adaptive Human Behavior and Physiology, 5*(2), 117-130.

Kordsmeyer, T. L., Freund, D., Vugt, M. V., & Penke, L. (2019). Honest signals of status: Facial and bodily dominance are related to success in physical but not nonphysical competition. *Evolutionary Psychology, 17*(3), 1474704919863164.

Kordsmeyer, T. L., Hunt, J., Puts, D. A., Ostner, J., & Penke, L. (2018). The relative importance of intra- and intersexual selection on human male sexually dimorphic traits. *Evolution and Human Behavior, 39*(4), 424-436.

Kordsmeyer, T. L., & Penke, L. (2017). The association of three indicators of developmental instability with mating success in humans. *Evolution and Human Behavior, 38*(6), 704-713.

Krams, I. A., Skrinda, I., Kecko, S., Moore, F. R., Krama, T., Kaasik, A., ... & Rantala, M. J. (2014). Body height affects the strength of immune response in young men, but not young women. *Scientific Reports*, *4*(1), 1-3.

Kuna, B., & Galbarczyk, A. (2018). Men with more masculine digit ratios are partnered with more attractive women. *Personality and Individual Differences, 124,* 8-11.

Kurzban, R., & Weeden, J. (2005). HurryDate: Mate preferences in action. *Evolution and Human Behavior, 26*(3), 227-244.

Krzyżanowska, M., Mascie-Taylor, C. N., & Thalabard, J. C. (2015). Is human mating for height associated with fertility? Results from a british national cohort study. *American Journal of Human Biology, 27*(4), 553-563.

Landolt, M. A., Lalumière, M. L., & Quinsey, V. L. (1995). Sex differences in intra-sex variations in human mating tactics: An evolutionary approach. *Ethology and Sociobiology*, *16*(1), 3-23.

Lassek, W. D., & Gaulin, S. J. (2008). Waist-hip ratio and cognitive ability: Is gluteofemoral fat a privileged store of neurodevelopmental resources?. *Evolution and Human Behavior*, *29*(1), 26-34.

Lassek, W. D., & Gaulin, S. J. (2009). Costs and benefits of fat-free muscle mass in men: Relationship to mating success, dietary requirements, and native immunity. *Evolution and Human Behavior, 30*(5), 322-328.

Lassek, W. D., & Gaulin, S. J. (2018a). Do the low WHRs and BMIs judged most attractive indicate better health?. *Evolutionary Psychology, 16*(4), 1474704918800063.

Lassek, W. D., & Gaulin, S. J. (2019). Evidence supporting nubility and reproductive value as the key to human female physical attractiveness. *Evolution and Human Behavior, 40*(5), 408-419.

Law Smith, M. J., Perrett, D. I., Jones, B. C., Cornwell, R. E., Moore, F. R., Feinberg, D. R., ... & Pitman, R. M. (2006). Facial appearance is a cue to oestrogen levels in women. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1583), 135-140.

Lee, A. J., Mitchem, D. G., Wright, M. J., Martin, N. G., Keller, M. C., & Zietsch, B. P. (2014). Genetic factors that increase male facial masculinity decrease facial attractiveness of female relatives. *Psychological Science*, *25*(2), 476-484.

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.

Lie, H. C., Rhodes, G., & Simmons, L. W. (2008). Genetic diversity revealed in human faces. *Evolution: International Journal of Organic Evolution, 62*(10), 2473-2486.

Lindenfors, P., Gittleman, J. L., & Jones, K. E. (2007). Sexual size dimorphism in mammals. *Sex, size and gender roles: Evolutionary studies of sexual size dimorphism, 16,* 26.

Lipson, S. F., & Ellison, P. T. (1996). EndocrinologyComparison of salivary steroid profiles in naturally occurring conception and non-conception cycles. *Human Reproduction, 11*(10), 2090-2096.

Little, A. C. (2015). Attraction and human mating. In *Evolutionary perspectives on social psychology* (pp. 319-332). Springer, Cham.

Little, A. C., Burriss, R. P., Jones, B. C., & Roberts, S. C. (2007). Facial appearance affects voting decisions. *Evolution and Human Behavior, 28*(1), 18-27.

Little, A. C., Burt, D. M., & Perrett, D. I. (2006). Assortative mating for perceived facial personality traits. *Personality and Individual Differences, 40*(5), 973-984.

Little, A. C., Penton-Voak, I. S., Burt, D. M., & Perrett, D. I. (2003). Investigating an imprinting-like phenomenon in humans: Partners and opposite-sex parents have similar hair and eye colour. *Evolution and Human Behavior, 24*(1), 43-51.

Little, A. C., Třebický, V., Havlíček, J., Roberts, S. C., & Kleisner, K. (2015). Human perception of fighting ability: Facial cues predict winners and losers in mixed martial arts fights. *Behavioral Ecology, 26*(6), 1470-1475.

Little, B. B., Malina, R. M., Buschang, P. H., & Little, L. R. (1989). Natural selection is not related to reduced body size in a rural subsistence agricultural community in southern Mexico. *Human Biology, 61*(2), 287-296.

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*, *9*(4), 20130049.

Longman, D. P., Surbey, M. K., Stock, J. T., & Wells, J. C. (2018). Tandem androgenic and psychological shifts in male reproductive effort following a manipulated "win" or "loss" in a sporting competition. *Human Nature, 29*(3), 283-310.

Lu, H., Shen, D., Wang, L., Niu, S., Bai, C., Ma, Z., & Huo, Z. (2017). Digit ratio (2D:4D) and handgrip strength are correlated in women (but not in men) in Hui ethnicity. *Early Human Development*, *109*, 21-25.

Lubahn, D. B., Moyer, J. S., Golding, T. S., Couse, J. F., Korach, K. S., & Smithies, O. (1993). Alteration of reproductive function but not prenatal sexual development after insertional disruption of the mouse estrogen receptor gene. *Proceedings of the National Academy of Sciences*, *90*(23), 11162-11166.

Luevano, V. X., Merkling, K., Sanchez-Pinelo, J. J., Shakeshaft, M., Velazquez, M. L. (2018, April). *Facial width-to-height ratio, psychopathy, sociosexual orientation, and accepting detection-risk in extra-pair relationships.* [Poster presentation]. The Western Psychological Association, Portland, OR, United States.

Lukas, W. D., Campbell, B. C., & Ellison, P. T. (2004). Testosterone, aging, and body composition in men from Harare, Zimbabwe. *American Journal of Human Biology: The Official Journal of the Human Biology Association, 16*(6), 704-712.

Lukaszewski, A. W., Larson, C. M., Gildersleeve, K. A., Roney, J. R., & Haselton, M. G. (2014). Condition-dependent calibration of men's uncommitted mating orientation: Evidence from multiple samples. *Evolution and Human Behavior, 35*(4), 319-326.

Lukaszewski, A. W., Simmons, Z. L., Anderson, C., & Roney, J. R. (2016). The role of physical formidability in human social status allocation. *Journal of Personality and Social Psychology, 110*(3), 385-406.

Luo, S., & Klohnen, E. C. (2005). Assortative mating and marital quality in newlyweds: a couple-centered approach. *Journal of Personality and Social Psychology, 88*(2), 304.

Maestripieri, D., Klimczuk, A., Traficonte, D., & Wilson, M. C. (2014). Ethnicity-related variation in sexual promiscuity, relationship status, and testosterone levels in men. *Evolutionary Behavioral Sciences, 8*(2), 96-108.

Manning, J. T. (2002). *Digit ratio: A pointer to fertility, behavior, and health.* Rutgers University Press.

Manning, J. T., Barley, L., Walton, J., Lewis-Jones, D. I., Trivers, R. L., Singh, D., ... & Szwed, A. (2000). The 2nd:4th digit ratio, sexual dimorphism, population differences, and reproductive success: Evidence for sexually antagonistic genes?. *Evolution and Human Behavior*, *21*(3), 163-183.

Manning, J. T., & Fink, B. (2008). Digit ratio (2D:4D), dominance, reproductive success, asymmetry, and sociosexuality in the BBC Internet Study. *American Journal of Human Biology, 20*(4), 451-461.

Manning, J. T., Henzi, P., Venkatramana, P., Martin, S., & Singh, D. (2003). Second to fourth digit ratio: Ethnic differences and family size in English, Indian and South African populations. *Annals of Human Biology, 30*(5), 579-588.

Manson, J. H., Wrangham, R. W., Boone, J. L., Chapais, B., Dunbar, R. I. M., Ember, C. R., ... & Worthman, C. M. (1991). Intergroup aggression in chimpanzees and humans [and comments and replies]. *Current Anthropology, 32*(4), 369-390.

Marcinkowska, U. M., Jasienska, G., & Prokop, P. (2018). A comparison of masculinity facial preference among naturally cycling, pregnant, lactating, and post-menopausal women. *Archives of Sexual Behavior*, *47*(5), 1367-1374.

Marcinkowska, U. M., Jones, B. C., & Lee, A. J. (2021). Self-rated attractiveness predicts preferences for sexually dimorphic facial characteristics in a culturally diverse sample. *Scientific Reports*, *11*(1), 1-8.

Marcinkowska, U. M., Rantala, M. J., Lee, A. J., Kozlov, M. V., Aavik, T., Cai, H., ... & Dixson, B. J. (2019). Women's preferences for men's facial masculinity are strongest under favorable ecological conditions. *Scientific Reports*, *9*(1), 1-10.

Marczak, M., Misiak, M., Sorokowska, A., & Sorokowski, P. (2018). No sex difference in digit ratios (2D:4D) in the traditional Yali of Papua and its meaning for the previous hypotheses on the inter-populational variability in 2D:4D. *American Journal of Human Biology*, *30*(2), e23078.

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.

Marlowe, F. (1998). The nubility hypothesis. *Human Nature*, *9*(3), 263-271.

Marti, B., Tuomilehto, J., Salomaa, V., Kartovaara, L., Korhonen, H. J., & Pietinen, P. (1991). Body fat distribution in the Finnish population: environmental determinants and predictive power for cardiovascular risk factor levels. *Journal of Epidemiology & Community Health*, *45*(2), 131-137.

Mazur, A., Halpern, C., & Udry, J. R. (1994). Dominant looking male teenagers copulate earlier. *Ethology and Sociobiology, 15*(2), 87-94.

McDade, T. W. (2003). Life history theory and the immune system: Steps toward a human ecological immunology. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *122*(S37), 100-125.

McIntyre, M. H., Chapman, J. F., Lipson, S. F., & Ellison, P. T. (2007). Index-to-ring finger length ratio (2D: 4D) predicts levels of salivary estradiol, but not progesterone, over the menstrual cycle. *American Journal of Human Biology: The Official Journal of the Human Biology Association*, *19*(3), 434-436.

McIntyre, M., Gangestad, S. W., Gray, P. B., Chapman, J. F., Burnham, T. C., O'Rourke, M. T., & Thornhill, R. (2006). Romantic involvement often reduces men’s testosterone levels--but not always: The moderating role of extrapair sexual interest. *Journal of Personality and Social Psychology, 91*(4), 642-651.

Međedović, J., & Bulut, T. (2019). A life-history perspective on body mass: Exploring the interplay between harsh environment, body mass, and mating success. *Evolutionary Behavioral Sciences, 13*(1), 84-92.

Mehta, P. H., & Prasad, S. (2015). The dual-hormone hypothesis: a brief review and future research agenda. *Current Opinion in Behavioral Sciences, 3,* 163-168.

Meins, E., Fernyhough, C., Arnott, B., Turner, M., & Leekam, S. R. (2011). Mother-versus infant-centered correlates of maternal mind-mindedness in the first year of life. *Infancy, 16*(2), 137-165.

Menken, J., Trussell, J., & Larsen, U. (1986). Age and infertility. *Science, 233*(4771), 1389-1394.

Mitani, J. C., Gros-Louis, J., & Richards, A. F. (1996). Sexual dimorphism, the operational sex ratio, and the intensity of male competition in polygynous primates. *The American Naturalist, 147*(6), 966-980.

Mitchem, D. G., Purkey, A. M., Grebe, N. M., Carey, G., Garver-Apgar, C. E., Bates, T. C., ... & Keller, M. C. (2014). Estimating the sex-specific effects of genes on facial attractiveness and sexual dimorphism. *Behavior Genetics*, *44*(3), 270-281.

Mitteroecker, P., Gunz, P., Windhager, S., & Schaefer, K. (2013). A brief review of shape, form, and allometry in geometric morphometrics, with applications to human facial morphology. *Hystrix, the Italian Journal of Mammalogy, 24*(1), 59-66.

Mitteroecker, P., Windhager, S., Müller, G. B., & Schaefer, K. (2015). The morphometrics of "masculinity" in human faces. *PLoS One, 10*(2), e0118374.

Monden, C. W., & Smits, J. (2009). Maternal height and child mortality in 42 developing countries. *American Journal of Human Biology: The Official Journal of the Human Biology Association*, *21*(3), 305-311.

Mondragón-Ceballos, R., Garcia Granados, M. D., Cerda-Molina, A. L., Chavira-Ramírez, R., & Hernández-López, L. E. (2015). Waist-to-hip ratio, but not body mass index, is associated with testosterone and estradiol concentrations in young women. *International Journal of Endocrinology, 2015*.

Montalcini, T., Ferro, Y., Salvati, M. A., Romeo, S., Miniero, R., & Pujia, A. (2016). Gender difference in handgrip strength of Italian children aged 9 to 10 years. *Italian Journal of Pediatrics, 42*(1), 1-6.

Mosing, M. A., Verweij, K. J., Madison, G., Pedersen, N. L., Zietsch, B. P., & Ullén, F. (2015). Did sexual selection shape human music? Testing predictions from the sexual selection hypothesis of music evolution using a large genetically informative sample of over 10,000 twins. *Evolution and Human Behavior, 36*(5), 359-366.

Muehlenbein, M. P., & Bribiescas, R. G. (2005). Testosterone-mediated immune functions and male life histories. *American Journal of Human Biology: The Official Journal of the Human Biology Association, 17*(5), 527-558.

Muller, M. N., Marlowe, F. W., Bugumba, R., & Ellison, P. T. (2009). Testosterone and paternal care in East African foragers and pastoralists. *Proceedings of the Royal Society B: Biological Sciences, 276*(1655), 347-354.

Muller, U., & Mazur, A. (1997). Facial dominance in Homo sapiens as honest signaling of male quality. *Behavioral Ecology, 8*(5), 569-579.

Nagelkerke, N. J., Bernsen, R. M., Sgaier, S. K., & Jha, P. (2006). Body mass index, sexual behaviour, and sexually transmitted infections: an analysis using the NHANES 1999-2000 data. *BMC Public Health, 6*(1), 1-9.

Neave, N., Laing, S., Fink, B., & Manning, J. T. (2003). Second to fourth digit ratio, testosterone and perceived male dominance. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 270*(1529), 2167-2172.

Nettle, D. (2002). Height and reproductive success in a cohort of British men. *Human Nature, 13*(4), 473-491.

Norgan, N. G. (1997). The beneficial effects of body fat and adipose tissue in humans. *International Journal of Obesity, 21*(9), 738-746.

Nowak, J., Pawłowski, B., Borkowska, B., Augustyniak, D., & Drulis-Kawa, Z. (2018). No evidence for the immunocompetence handicap hypothesis in male humans. *Scientific Reports, 8*(1), 1-11.

Olsen, N. J., & Kovacs, W. J. (1996). Gonadal steroids and immunity. *Endocrine Reviews, 17*(4), 369-384.

Parker, G. A. (1974). Assessment strategy and the evolution of fighting behaviour. *Journal of Theoretical Biology, 47*(1), 223-243.

Pawlowski, B., G Boothroyd, L., I Perrett, D., & Kluska, S. (2008). Is female attractiveness related to final reproductive success?. *Collegium Antropologicum, 32*(2), 457-460.

Pawlowski, B., Borkowska, B., Nowak, J., Augustyniak, D., & Drulis-Kawa, Z. (2018). Human body symmetry and immune efficacy in healthy adults. *American Journal of Physical Anthropology, 167*(2), 207-216.

Pawlowski, B., Dunbar, R. I., & Lipowicz, A. (2000). Tall men have more reproductive success. *Nature, 403*(6766), 156-156.

Pawlowski, B., Nowak, J., Borkowska, B. A. R. B. A. R. A., & Drulis-Kawa, Z. (2014). Human body morphology, prevalence of nasopharyngeal potential bacterial pathogens, and immunocompetence handicap principal. *American Journal of Human Biology, 26*(3), 305-310.

Penke, L., & Asendorpf, J. B. (2008). Beyond global sociosexual orientations: a more differentiated look at sociosexuality and its effects on courtship and romantic relationships. *Journal of Personality and Social Psychology, 95*(5), 1113-1135.

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*(4), 229-241.

Penton-Voak, I. S., & Perrett, D. I. (2000). Female preference for male faces changes cyclically: Further evidence. *Evolution and Human Behavior, 21*(1), 39-48.

Pereira, K. J., da Silva, C. S. A., Havlíček, J., Kleisner, K., Varella, M. A. C., Pavlovič, O., & Valentova, J. V. (2019). Femininity-masculinity and attractiveness - Associations between self-ratings, third-party ratings and objective measures. *Personality and Individual Differences, 147*, 166-171.

Perrett, D. I., May, K. A., & Yoshikawa, S. (1994). Facial shape and judgements of female attractiveness. *Nature, 368*(6468), 239-242.

Pérusse, D. (1993). Cultural and reproductive success in industrial societies: Testing the relationship at the proximate and ultimate levels. *Behavioral and Brain Sciences, 16*(2), 267-283.

Peters, M., Simmons, L. W., & Rhodes, G. (2008). Testosterone is associated with mating success but not attractiveness or masculinity in human males. *Animal Behaviour, 76*(2), 297-303.

Petersen, A. C., Crockett, L., Richards, M., & Boxer, A. (1988). A self-report measure of pubertal status: Reliability, validity, and initial norms. *Journal of Youth and Adolescence, 17*(2), 117-133.

Peterson, R. A., & Brown, S. P. (2005). On the use of beta coefficients in meta-analysis. *Journal of Applied Psychology, 90*(1), 175-181.

Pflüger, L. S., Oberzaucher, E., Katina, S., Holzleitner, I. J., & Grammer, K. (2012). Cues to fertility: Perceived attractiveness and facial shape predict reproductive success. *Evolution and Human Behavior, 33*(6), 708-714.

Phalane, K. G., Tribe, C., Steel, H. C., Cholo, M. C., & Coetzee, V. (2017). Facial appearance reveals immunity in African men. *Scientific Reports*, *7*(1), 1-9.

Pisanski, K., Fraccaro, P. J., Tigue, C. C., O’Connor, J. J., Röder, S., Andrews, P. W., ... & Feinberg, D. R. (2014). Vocal indicators of body size in men and women: A meta-analysis. *Animal Behaviour, 95*, 89-99.

Plavcan, J. M. (2001). Sexual dimorphism in primate evolution. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *116*(S33), 25-53.

Plavcan, J. M. (2011). Understanding dimorphism as a function of changes in male and female traits. *Evolutionary Anthropology: Issues, news, and reviews, 20*(4), 143-155.

Plavcan, J. M. (2012). Sexual size dimorphism, canine dimorphism, and male-male competition in primates. *Human Nature*, *23*(1), 45-67.

Pollet, T. V., & Saxton, T. K. (2019). How diverse are the samples used in the journals ‘evolution & human behavior’ and ‘evolutionary psychology’?. *Evolutionary Psychological Science, 5*(3), 357-368.

Pollet, T. V., van der Meij, L., Cobey, K. D., & Buunk, A. P. (2011). Testosterone levels and their associations with lifetime number of opposite sex partners and remarriage in a large sample of American elderly men and women. *Hormones and Behavior*, *60*(1), 72-77.

Polo, P., Muñoz-Reyes, J. A., Pita, M., Shackelford, T. K., & Fink, B. (2019). Testosterone dependent facial and body traits predict men’s sociosexual attitudes and behaviours. *American Journal of Human Biology, 31*(3), e23235.

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*(1654), 153-159.

Pradhan, A., & Olsson, P. E. (2020). Sex differences in severity and mortality from COVID-19: Are males more vulnerable?. *Biology of Sex Differences, 11*(1), 1-11.

Price, M. E., Pound, N., Dunn, J., Hopkins, S., & Kang, J. (2013). Body shape preferences: Associations with rater body shape and sociosexuality. *PloS One, 8*(1), e52532.

Probst, F., Bobst, C., & Lobmaier, J. S. (2016). Testosterone-to-oestradiol ratio is associated with female facial attractiveness. *Quarterly Journal of Experimental Psychology, 69*(1), 89-99.

Prokop, P., & Fedor, P. (2011). Physical attractiveness influences reproductive success of modern men. *Journal of Ethology, 29*(3), 453-458.

Prokop, P., & Fedor, P. (2013). Associations between body morphology, mating success and mate preferences among Slovak males and females. *Anthropologischer Anzeiger, 70*(2), 121-135.

Prokop, P., Obertová, Z., & Fedor, P. (2010). Paternity cues and mating opportunities: what makes fathers good?. *Acta Ethologica*, *13*(2), 101-107.

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

Puts, D. (2016). Human sexual selection. *Current Opinion in Psychology, 7,* 28-32.

Puts, D. A., Apicella, C. L., & Cárdenas, R. A. (2012). Masculine voices signal men’s threat potential in forager and industrial societies. *Proceedings of the Royal Society B: Biological Sciences, 279*(1728), 601-609.

Puts, D. A., & Aung, T. (2019). Does men’s voice pitch signal formidability? A reply to Feinberg et al. *Trends in Ecology & Evolution, 34*(3), 189-190.

Puts, D. A., Gaulin, S. J., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior, 27*(4), 283-296.

Puts, D. A., Pope, L. E., Hill, A. K., Cárdenas, R. A., Welling, L. L., Wheatley, J. R., & Breedlove, S. M. (2015). Fulfilling desire: Evidence for negative feedback between men’s testosterone, sociosexual psychology, and sexual partner number. *Hormones and Behavior, 70,* 14-21.

Putz, D. A., Gaulin, S. J., Sporter, R. J., & McBurney, D. H. (2004). Sex hormones and finger length: What does 2D:4D indicate?. *Evolution and Human Behavior, 25*(3), 182-199.

R Development Core Team (2019; 2021). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna*.

Rahman, Q., Korhonen, M., & Aslam, A. (2005). Sexually dimorphic 2D: 4D ratio, height, weight, and their relation to number of sexual partners. *Personality and Individual Differences, 39*(1), 83-92.

Raine, J., Pisanski, K., Oleszkiewicz, A., Simner, J., & Reby, D. (2018). Human listeners can accurately judge strength and height relative to self from aggressive roars and speech. *Iscience, 4*, 273-280.

Rantala, M. J., Coetzee, V., Moore, F. R., Skrinda, I., Kecko, S., Krama, T., ... & Krams, I. (2013a). Adiposity, compared with masculinity, serves as a more valid cue to immunocompetence in human mate choice. *Proceedings of the Royal Society B: Biological Sciences, 280*(1751), 20122495.

Rantala, M. J., Coetzee, V., Moore, F. R., Skrinda, I., Kecko, S., Krama, T., ... & Krams, I. (2013b). Facial attractiveness is related to women's cortisol and body fat, but not with immune responsiveness. *Biology Letters*, *9*(4), 20130255.

Rantala, M. J., Moore, F. R., Skrinda, I., Krama, T., Kivleniece, I., Kecko, S., & Krams, I. (2012). Evidence for the stress-linked immunocompetence handicap hypothesis in humans. *Nature Communications*, *3*(1), 1-5.

Reed, T., Fabsitz, R. R., Selby, J. V., & Carmelli, D. (1991). Genetic influences and grip strength norms in the NHLBI twin study males aged 59-69. *Annals of Human Biology*, *18*(5), 425-432.

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology*, *57*, 199-226.

Rhodes, G., Chan, J., Zebrowitz, L. A., & Simmons, L. W. (2003). Does sexual dimorphism in human faces signal health?. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(suppl_1), S93-S95.

Rhodes, G., Simmons, L. W., & Peters, M. (2005). Attractiveness and sexual behavior: Does attractiveness enhance mating success?. *Evolution and Human Behavior*, *26*(2), 186-201.

Richards, G., Baron-Cohen, S., van Steen, T., & Galvin, J. (2020). Assortative mating and digit ratio (2D:4D): A pre-registered empirical study and meta-analysis. *Early Human Development*, *151*, 105159.

Richards, G., Gomes, M., & Ventura, T. (2019). Testosterone measured from amniotic fluid and maternal plasma shows no significant association with directional asymmetry in newborn digit ratio (2D:4D). *Journal of Developmental Origins of Health and Disease, 10*(3), 362-367.

Rilling, J. K., Kaufman, T. L., Smith, E. O., Patel, R., & Worthman, C. M. (2009). Abdominal depth and waist circumference as influential determinants of human female attractiveness. *Evolution and Human Behavior, 30*(1), 21-31.

Roberts, M. L., Buchanan, K. L., & Evans, M. R. (2004). Testing the immunocompetence handicap hypothesis: A review of the evidence. *Animal Behaviour, 68*(2), 227-239.

Roberts, S. C., Little, A. C., Gosling, L. M., Perrett, D. I., Carter, V., Jones, B. C., ... & Petrie, M. (2005). MHC-heterozygosity and human facial attractiveness. *Evolution and Human Behavior, 26*(3), 213-226.

Robinson, M. R., Kleinman, A., Graff, M., Vinkhuyzen, A. A., Couper, D., Miller, M. B., ... & Visscher, P. M. (2017). Genetic evidence of assortative mating in humans. *Nature Human Behaviour, 1*(1), 1-13.

Rosenfield, K. A., Sorokowska, A., Sorokowski, P., & Puts, D. A. (2020). Sexual selection for low male voice pitch among Amazonian forager-horticulturists. *Evolution and Human Behavior, 41*(1), 3-11.

Roved, J., Westerdahl, H., & Hasselquist, D. (2017). Sex differences in immune responses: Hormonal effects, antagonistic selection, and evolutionary consequences. *Hormones and Behavior, 88*, 95-105.

Rule, N. O., & Ambady, N. (2008). The face of success: Inferences from chief executive officers' appearance predict company profits. *Psychological Science, 19*(2), 109-111.

Salonia, A., Pontillo, M., Capogrosso, P., Gregori, S., Tassara, M., Boeri, L., ... & Montorsi, F. (2021). Severely low testosterone in males with COVID-19: A case-control study. *Andrology, 9*(4), 1043-1052.

Samal, A., Subramani, V., & Marx, D. (2007). Analysis of sexual dimorphism in human face. *Journal of Visual Communication and Image Representation, 18*(6), 453-463.

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.

Schaefer, K., Mitteroecker, P., Gunz, P., Bernhard, M., & Bookstein, F. L. (2004). Craniofacial sexual dimorphism patterns and allometry among extant hominids. *Annals of Anatomy-Anatomischer Anzeiger, 186*(5-6), 471-478.

Schild, C., Stern, J., Penke, L., & Zettler, I. (2020). Voice pitch - A valid indicator of one's unfaithfulness in committed relationships?. *Adaptive Human Behavior and Physiology,* 1-16.

Schwarz, S., Mustafić, M., Hassebrauck, M., & Jörg, J. (2011). Short-and long-term relationship orientation and 2D:4D finger-length ratio. *Archives of Sexual Behavior, 40*(3), 565-574.

Scott, E. C., & Bajema, C. J. (1982). Height, weight and fertility among the participants of the third Harvard growth study. *Human Biology, 54*(3), 501-516.

Scott, I. M., Clark, A. P., Boothroyd, L. G., & Penton-Voak, I. S. (2013). Do men's faces really signal heritable immunocompetence?. *Behavioral Ecology, 24*(3), 579-589.

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, 111*(40), 14388-14393.

Scott, I. M., 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.

Sear, R., & Coall, D. (2011). How much does family matter? Cooperative breeding and the demographic transition. *Population and Development Review*, *37,* 81-112.

Sear, R., & Marlowe, F. W. (2009). How universal are human mate choices? Size does not matter when Hadza foragers are choosing a mate. *Biology Letters, 5*(5), 606-609.

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., & Pound, N. (2012). The importance of physical strength to human males. *Human Nature*, *23*(1), 30-44.

Sell, A., Lukazsweski, A. W., & Townsley, M. (2017). Cues of upper body strength account for most of the variance in men's bodily attractiveness. *Proceedings of the Royal Society B: Biological Sciences, 284*(1869), 20171819.

Shackelford, T. K., & Larsen, R. J. (1999). Facial attractiveness and physical health. *Evolution and Human Behavior*, *20*(1), 71-76.

Shoup, M. L., & Gallup, G. G. (2008). Men's faces convey information about their bodies and their behavior: What you see is what you get. *Evolutionary Psychology, 6*(3), 147470490800600311.

Silva, A. S., Lummaa, V., Muller, U., Raymond, M., & Alvergne, A. (2012). Facial attractiveness and fertility in populations with low levels of modern birth control. *Evolution and Human Behavior*, *33*(5), 491-498.

Silventoinen, K. (2003). Determinants of variation in adult body height. *Journal of Biosocial Science*, *35*(2), 263-285.

Silventoinen, K., Kaprio, J., Lahelma, E., Viken, R. J., & Rose, R. J. (2003). Assortative mating by body height and BMI: Finnish twins and their spouses. *American Journal of Human Biology, 15*(5), 620-627.

Sim, K. (2013). The relationship between sex-typical body shape and quality indicators. *Journal of Social, Evolutionary, and Cultural Psychology, 7*(2), 97-120.

Sim, K., & Chun, W. Y. (2016). The relationships of assertiveness and responsiveness to sexual behavior. *Psychologia*, *59*(1), 50-69.

Simmons, Z. L., & Roney, J. R. (2011). Variation in CAG repeat length of the androgen receptor gene predicts variables associated with intrasexual competitiveness in human males. *Hormones and Behavior, 60*(3), 306-312.

Simpson, J. A., & Gangestad, S. W. (1991). Individual differences in sociosexuality: evidence for convergent and discriminant validity. *Journal of Personality and Social psychology, 60*(6), 870-883.

Singh, D. (1993). Adaptive significance of female physical attractiveness: Role of waist-to hip ratio. *Journal of Personality and Social Psychology, 65*(2), 293-307.

Singh, D. (1994). Ideal female body shape: Role of body weight and waist-to-hip ratio. *International Journal of Eating Disorders, 16*(3), 283-288.

Skrinda, I., Krama, T., Kecko, S., Moore, F. R., Kaasik, A., Meija, L., ... & Krams, I. (2014). Body height, immunity, facial and vocal attractiveness in young men. *Naturwissenschaften*, *101*(12), 1017-1025.

Slatcher, R. B., Mehta, P. H., & Josephs, R. A. (2011). Testosterone and self-reported dominance interact to influence human mating behavior. *Social Psychological and Personality Science*, *2*(5), 531-539.

Smith, R. J., & Jungers, W. L. (1997). Body mass in comparative primatology. *Journal of Human Evolution*, *32*(6), 523-559.

Smith, J. M., & Maynard-Smith, J. (1978). *The evolution of sex* (Vol. 4). Cambridge: Cambridge University Press.

Smith, K. M., Olkhov, Y. M., Puts, D. A., & Apicella, C. L. (2017). Hadza men with lower voice pitch have a better hunting reputation. *Evolutionary Psychology, 15*(4), 1474704917740466.

Sneade, M., & Furnham, A. (2016). Hand grip strength and self-perceptions of physical attractiveness and psychological well-being. *Evolutionary Psychological Science, 2*(2), 123-128.

Snodell, S. F., Nanda, R. S., & Currier, G. F. (1993). A longitudinal cephalometric study of transverse and vertical craniofacial growth. *American Journal of Orthodontics and Dentofacial Orthopedics, 104*(5), 471-483.

Sorokowski, P., Sorokowska, A., & Danel, D. P. (2013). Why pigs are important in Papua? Wealth, height and reproductive success among the Yali tribe of West Papua. *Economics & Human Biology, 11*(3), 382-390.

Spisak, B. R., Dekker, P. H., Krüger, M., & Van Vugt, M. (2012). Warriors and peacekeepers: Testing a biosocial implicit leadership hypothesis of intergroup relations using masculine and feminine faces. *PloS One, 7*(1), e30399.

Stefan, N., Häring, H. U., Hu, F. B., & Schulze, M. B. (2016). Divergent associations of height with cardiometabolic disease and cancer: Epidemiology, pathophysiology, and global implications. *The Lancet Diabetes & Endocrinology, 4*(5), 457-467.

Steiner, E. T. (2011). Testosterone and vasopressin in men's reproductive behavior. [Doctoral dissertation, University of Oregon].

Štěrbová, Z., & Valentova, J. (2012). Influence of homogamy, complementarity, and sexual imprinting on mate choice. *Anthropologie (1962-), 50*(1), 47-60.

Stern, J., Karastoyanova, K., Kandrik, M., Torrance, J., Hahn, A. C., Holzleitner, I., ... & Jones, B. C. (2020). Are sexual desire and sociosexual orientation related to men's salivary steroid hormones?. *Adaptive Human Behavior and Physiology, 6*(4), 447-466.

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64*(3), 479-498.

Storey, J. D., Bass, A. J., Alan Dabney, A., & Robinson, D. (2017). qvalue: Q-value estimation for false discovery rate control. R package version 2.15.0. http://github.com/StoreyLab/qvalue

Strong, H. L. (2014). *Association of prenatal androgen exposure with attention and sociosexual orientation* [Doctoral dissertation, California State University].

Strong, H., & Luevano, V. X. (2014, May). *Prenatal androgen exposure predicts relationship type preference but not experience* [Poster presentation]. The Association for Psychological Science, San Diego, CA, United States.

Stulp, G., Buunk, A. P., Pollet, T. V., Nettle, D., & Verhulst, S. (2013). Are human mating preferences with respect to height reflected in actual pairings?. *PLoS One, 8*(1), e54186.

Stulp, G., Buunk, A. P., & Pollet, T. V. (2013). Women want taller men more than men want shorter women. *Personality and Individual Differences*, *54*(8), 877-883.

Stulp, G., Simons, M. J., Grasman, S., & Pollet, T. V. (2017). Assortative mating for human height: A meta-analysis. *American Journal of Human Biology, 29*(1), e22917.

Subramanian, S. V., Ackerson, L. K., Smith, G. D., & John, N. A. (2009). Association of maternal height with child mortality, anthropometric failure, and anemia in India. *Jama, 301*(16), 1691-1701.

Suire, A., Raymond, M., & Barkat-Defradas, M. (2018). Human vocal behavior within competitive and courtship contexts and its relation to mating success. *Evolution and Human Behavior, 39*(6), 684-691.

Suire, A., Raymond, M., & Barkat-Defradas, M. (2019). Male vocal quality and its relation to females' preferences. *Evolutionary Psychology, 17*(3), 1474704919874675.

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*(1507), 2285-2289.

Tao, H. L., & Yin, C. C. (2016). Height, partners and offspring: Evidence from Taiwan. *Journal of Biosocial Science, 48*(5), 593-615.

Tea-Makorn, P. P., & Kosinski, M. (2020). Spouses' faces are similar but do not become more similar with time. *Scientific Reports, 10*(1), 1-6.

The World Bank. (2018). Fertility rate, total (births per woman). https://data.worldbank.org/indicator/SP.DYN.TFRT.IN

Thiessen, D., & Gregg, B. (1980). Human assortative mating and genetic equilibrium: An evolutionary perspective. *Ethology and Sociobiology, 1*(2), 111-140.

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.

Thornborrow, T., Jucker, J. L., Boothroyd, L. G., & Tovée, M. J. (2018). Investigating the link between television viewing and men's preferences for female body size and shape in rural Nicaragua. *Evolution and Human Behavior, 39*(5), 538-546.

Thornhill, R., & Gangestad, S. W. (2006). Facial sexual dimorphism, developmental stability, and susceptibility to disease in men and women. *Evolution and Human Behavior, 27*(2), 131-144.

Timmons, J. A. (2011). Variability in training-induced skeletal muscle adaptation. *Journal of Applied Physiology, 110*(3), 846-853.

Titze, I. R. (2000). *Principles of voice production*. Iowa City, IA7 National Center for Voice and Speech.

Tovée, M. J., Maisey, D. S., Vale, E. L., & Cornelissen, P. L. (1999). Characteristics of male attractiveness for women. *The Lancet, 353*(9163), 1500.

Traish, A. M., & Morgentaler, A. (2021). What’s testosterone got to do with it? A critical assessment of the contribution of testosterone to gender disparities in COVID-19 infections and deaths. *Androgens: Clinical Research and Therapeutics, 2*(1), 18-35.

Třebický, V., Fialová, J., Stella, D., Coufalová, K., Pavelka, R., Kleisner, K., ... & Havlíček, J. (2019). Predictors of fighting ability inferences based on faces. *Frontiers in Psychology, 9,* 2740.

Trumble, B. C., Blackwell, A. D., Stieglitz, J., Thompson, M. E., Suarez, I. M., Kaplan, H., & Gurven, M. (2016). Associations between male testosterone and immune function in a pathogenically stressed forager-horticultural population. *American Journal of Physical Anthropology, 161*(3), 494-505.

van Anders, S. M. (2010). Gonadal steroids and salivary IgA in healthy young women and men. *American Journal of Human Biology, 22*(3), 348-352.

van Anders, S. M., & Hampson, E. (2005). Waist-to-hip ratio is positively associated with bioavailable testosterone but negatively associated with sexual desire in healthy premenopausal women. *Psychosomatic Medicine, 67*(2), 246-250.

van Anders, S. M., Hamilton, L. D., & Watson, N. V. (2007). Multiple partners are associated with higher testosterone in North American men and women. *Hormones and Behavior, 51*(3), 454-459.

van Bokhoven, I., Van Goozen, S. H., Van Engeland, H., Schaal, B., Arseneault, L., Séguin, J. R., ... & Tremblay, R. E. (2006). Salivary testosterone and aggression, delinquency, and social dominance in a population-based longitudinal study of adolescent males. *Hormones and Behavior, 50*(1), 118-125.

Van Dongen, S. (2014). Associations among facial masculinity, physical strength, fluctuating asymmetry and attractiveness in young men and women. *Annals of Human Biology, 41*(3), 205-213.

Van Dongen, S., & Gangestad, S. W. (2011). Human fluctuating asymmetry in relation to health and quality: A meta-analysis. *Evolution and Human Behavior, 32*(6), 380-398.

Van Dongen, S., & Sprengers, E. (2012). Hand grip strength in relation to morphological measures of masculinity, fluctuating asymmetry and sexual behaviour in males and females. *Sex Hormones. InTech*, 293-306.

Varella, M. A. C., Valentova, J. V., Pereira, K. J., & Bussab, V. S. R. (2014). Promiscuity is related to masculine and feminine body traits in both men and women: Evidence from Brazilian and Czech samples. *Behavioural Processes, 109*, 34-39.

Vehmas, T., Solovieva, S., & Leino-Arjas, P. (2006). Radiographic 2D: 4D index in females: no relation to anthropometric, behavioural, nutritional, health-related, occupational or fertility variables. *Journal of Negative Results in BioMedicine, 5*(1), 1-7.

Venners, S. A., Liu, X., Perry, M. J., Korrick, S. A., Li, Z., Yang, F., ... & Wang, X. (2006). Urinary estrogen and progesterone metabolite concentrations in menstrual cycles of fertile women with non-conception, early pregnancy loss or clinical pregnancy. *Human Reproduction, 21*(9), 2272-2280.

Ventura, T., Gomes, M. C., Pita, A., Neto, M. T., & Taylor, A. (2012). Digit ratio (2D:4D) in newborns: Influences of prenatal testosterone and maternal environment. *Early Human Development, 89*(2), 107-112.

Verdonck, A., Gaethofs, M., Carels, C., & de Zegher, F. (1999). Effect of low-dose testosterone treatment on craniofacial growth in boys with delayed puberty. *The European Journal of Orthodontics, 21*(2), 137-143.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1-48.

von Rueden, C., Gurven, M., & Kaplan, H. (2008). The multiple dimensions of male social status in an Amazonian society. *Evolution and Human Behavior, 29*(6), 402-415.

von Rueden, C., Gurven, M., & Kaplan, H. (2011). Why do men seek status? Fitness payoffs to dominance and prestige. P*roceedings of the Royal Society B: Biological Sciences, 278*(1715), 2223-2232.

von Rueden, C. R., & Jaeggi, A. V. (2016). Men's status and reproductive success in 33 nonindustrial societies: Effects of subsistence, marriage system, and reproductive strategy. *Proceedings of the National Academy of Sciences, 113*(39), 10824-10829.

Voracek, M., Dressler, S. G., & Manning, J. T. (2007). Evidence for assortative mating on digit ratio (2D:4D), a biomarker for prenatal androgen exposure. *Journal of Biosocial Science, 39*(4), 599-612.

Voracek, M., Pum, U., & Dressler, S. G. (2010). Investigating digit ratio (2D:4D) in a highly male-dominated occupation: The case of firefighters. *Scandinavian Journal of Psychology, 51*(2), 146-156.

Vukovic, J., Boothroyd, L. G., Meins, E., & Burt, D. M. (2015). Concurrent parent–child relationship quality is associated with an imprinting-like effect in children's facial preferences. *Evolution and Human Behavior*, *36*(4), 331-336.

Vukovic, J., Jones, B. C., DeBruine, L., Feinberg, D. R., Smith, F. G., Little, A. C., ... & Main, J. (2010). Women's own voice pitch predicts their preferences for masculinity in men's voices. *Behavioral Ecology, 21*(4), 767-772.

Walker, P. L. (2001). A bioarchaeological perspective on the history of violence. *Annual Review of Anthropology, 30*(1), 573-596.

Walker, R. S., Hill, K. R., Flinn, M. V., & Ellsworth, R. M. (2011). Evolutionary history of hunter-gatherer marriage practices. *PloS One, 6*(4), e19066.

Walther, A., Philipp, M., Lozza, N., & Ehlert, U. (2016). The rate of change in declining steroid hormones: A new parameter of healthy aging in men?. *Oncotarget, 7*(38), 60844.

Walther, A., Mahler, F., Debelak, R., & Ehlert, U. (2017a). Psychobiological protective factors modifying the association between age and sexual health in men: Findings from the men's health 40+ study. *American Journal of Men's health, 11(*3), 737-747.

Walther, A., Philipp, M., Lozza, N., & Ehlert, U. (2017b). Emotional support, depressive symptoms, and age-related alterations in male body composition: Cross-sectional findings from the men's health 40+ study. *Frontiers in Psychology, 8,* 1075.

Walther, A., Waldvogel, P., Noser, E., Ruppen, J., & Ehlert, U. (2017c). Emotions and steroid secretion in aging men: A multi-study report. *Frontiers in Psychology, 8,* 1722.

Waynforth, D. (1998). Fluctuating asymmetry and human male life-history traits in rural Belize. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 265*(1405), 1497-1501.

Weatherhead, P. J., & Robertson, R. J. (1979). Offspring quality and the polygyny threshold: "The sexy son hypothesis". *The American Naturalist, 113*(2), 201-208.

Weeden, J., & Sabini, J. (2005). Physical attractiveness and health in Western societies: A review. *Psychological Bulletin, 131*(5), 635.

Weeden, J., & Sabini, J. (2007). Subjective and objective measures of attractiveness and their relation to sexual behavior and sexual attitudes in university students. *Archives of Sexual Behavior, 36*(1), 79-88.

Wells, J. C. (2007). Sexual dimorphism of body composition. *Best practice & research Clinical endocrinology & metabolism, 21*(3), 415-430.

Wells, J. C., Griffin, L., & Treleaven, P. (2010). Independent changes in female body shape with parity and age: A life-history approach to female adiposity. *American Journal of Human Biology: The Official Journal of the Human Biology Association*, *22*(4), 456-462.

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., 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), 20151351.

Wind, A. E., Takken, T., Helders, P. J., & Engelbert, R. H. (2010). Is grip strength a predictor for total muscle strength in healthy children, adolescents, and young adults?. *European Journal of Pediatrics, 169*(3), 281-287.

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.

Winkler, E. M., & Kirchengast, S. (1994). Body dimensions and differential fertility in! Kung San males from Namibia. *American Journal of Human Biology, 6*(2), 203-213.

Witkower, Z., & Tracy, J. L. (2019). A facial-action imposter: How head tilt influences perceptions of dominance from a neutral face. *Psychological Science, 30*(6), 893-906.

Wood, W., Kressel, L., Joshi, P. D., & Louie, B. (2014). Meta-analysis of menstrual cycle effects on women's mate preferences. *Emotion Review, 6*(3), 229-249.

Xu, Y., Norton, S., & Rahman, Q. (2018). Early life conditions, reproductive and sexuality related life history outcomes among human males: A systematic review and meta-analysis. *Evolution and Human Behavior*, *39*(1), 40-51.

Younis, J. S., Skorecki, K., & Abassi, Z. (2021). The double edge sword of testosterone's role in the COVID-19 pandemic. *Frontiers in Endocrinology, 12*, 607179.

Zahavi, A. (1975). Mate selection—a selection for a handicap. *Journal of theoretical Biology, 53*(1), 205-214.

Zahavi, A., & Zahavi, A. (1999). *The handicap principle: A missing piece of Darwin's puzzle.* Oxford University Press.

Zaidi, A. A., White, J. D., Mattern, B. C., Liebowitz, C. R., Puts, D. A., Claes, P., & Shriver, M. D. (2019). Facial masculinity does not appear to be a condition-dependent male ornament and does not reflect MHC heterozygosity in humans. *Proceedings of the National Academy of Sciences, 116*(5), 1633-1638.

Zajonc, R. B., Adelmann, P. K., Murphy, S. T., & Niedenthal, P. M. (1987). Convergence in the physical appearance of spouses. *Motivation and Emotion, 11*(4), 335-346.

Zandman-Goddard, G., Peeva, E., & Shoenfeld, Y. (2007). Gender and autoimmunity. *Autoimmunity Reviews*, *6*(6), 366-372.

Zebrowitz, L. A., & Rhodes, G. (2004). Sensitivity to "bad genes" and the anomalous face overgeneralization effect: Cue validity, cue utilization, and accuracy in judging intelligence and health. *Journal of Nonverbal Behavior, 28*(3), 167-185.

Żelaźniewicz, A., Nowak, J., Łącka, P., & Pawłowski, B. (2020). Facial appearance and metabolic health biomarkers in women. *Scientific Reports, 10*(1), 1-8.

Zheng, Z., & Cohn, M. J. (2011). Developmental basis of sexually dimorphic digit ratios. *Proceedings of the National Academy of Sciences, 108*(39), 16289-16294.

Zheng, W., Suzuki, K., Yokomichi, H., Sato, M., & Yamagata, Z. (2013). Multilevel longitudinal analysis of sex differences in height gain and growth rate changes in Japanese school-aged children. *Journal of Epidemiology, 23*(4), 275-279.

Özaltin, E., Hill, K., & Subramanian, S. V. (2010). Association of maternal stature with offspring mortality, underweight, and stunting in low-to middle-income countries. *Jama*, *303*(15), 1507-1516.

HESA. (2021, February). *Table T1 – Participation of under-represented groups in higher education 2015/16 to 2019/20*. https://www.hesa.ac.uk/data-and-analysis/performance-indicators/widening-participation/table-t1