INDIVIDUAL VARIATION IN HUMAN MATE CHOICE

Thesis submitted in accordance with the
requirements of the University of Liverpool for the
degree of Doctor in Philosophy

by

Tamsin K. Saxton

August 2008

ABSTRACT

People vary in mate choice preferences and behaviour, and these individual differences may be adaptive. Such differences are considered across three different time scales: the short-term, with reference to olfactory stimuli; the medium-term, with reference to the menstrual cycle; and the long-term, with reference to the development of attractiveness judgments during adolescence.

The olfactory stimulus androstadienone, a chemical constituent of sweat and purported human pheromone, is shown to enhance women's ratings of men's attractiveness in the context of a speed-dating event. Some evidence is also given for the influence of androstadienone on women's perceptions of men's attractiveness rated in photographs and odour samples. Interactions between androstadienone-linked attractiveness enhancement and male mate quality gauged in terms of attractiveness and MHC heterozygosity are also examined.

The high-fertility phase of the menstrual cycle is shown to correspond to increases in women's attractiveness in a number of different modalities. Unexpectedly, raters are not better at selecting a recording of a woman as more attractive at high-fertility when they have access to visual, vocal and dynamic information compared with more limited single-channel information. Attractiveness changes may be more salient in the face, and irrespective of the menstrual cycle phase, facial attractiveness, followed by body and speech attractiveness, are shown to predict ratings of holistic individual attractiveness. Next, women's judgments of men's facial attractiveness are shown to trade off absolute and relative quality in a hierarchical, context-dependent manner, privileging absolute quality for short-term relationships and relative quality for long-term relationships.

Adolescents' attractiveness judgments of purported cues of mate quality in same-age stimuli are examined in a longitudinal study. Increased age is associated with higher ratings of facial symmetry, averageness and femininity, with higher ratings of low pitch in boys' voices, and with lower ratings of high pitch in girls' voices. This pattern is generally supportive of the hypothesis that children's perceptions of genetic quality emerge as they become capable of reproducing. Pubertal development itself controlling for age is also shown to have significant effects on attractiveness judgments, although these demonstrate a mixed pattern. The finding that even the youngest children in the sample rated facial symmetry, averageness and femininity as attractive suggests that these cues may provide non-sexual, social information. Increased or decreased exposure to same- and opposite-sex same-age children, by virtue of attendance at a mixed-sex compared with single-sex school, or by the presence of siblings, is shown to have mixed effects on facial and vocal attractiveness judgments.

Finally, results are discussed in relation to the evolutionary psychology research paradigm. It is argued that functionally-motivated research assuming adaptive variation across different time scales is productive.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

With thanks to...

... my colleagues at Liverpool University, including the evolutionary psychology group at Liverpool University, and in particular Katherine Andrews, Caroline Betteridge, Nassima Boulkroune, Robert Burriss, Ines Canavarro, Claire Cunningham, Robin Dunbar, Camille Ferdenzi, Chris Hassall, Kostas Kaskatis, Shelly Kemp, Austen Krill, John Lycett, Minna Lyons, Marc Mehu, Augusto Montiel-Castro, Steve Platek, Sam Roberts, Hannah Rowland, Yvan Russell, Suzanne Schultz, Maria Stylianou and Alex Wall.

... the group at Charles University who kindly hosted me for three months, headed up by Jan Havlíček and including Dagmar Kohoutová, Anna Kotrčová, Pavlína Lenochová, Věra Pivoňková and Jaroslava Valentová.

...staff at the Psychology department of Liverpool University who allowed me to join in and carry out research activities there, including Valerie Adams, Bill Bimson and Penny Lewis.

... the University of Liverpool for my Ph.D. studentship, the Owen Aldis scholarship for research and travel funding, and the ASAB conference travel fund.

....my supervisors, Craig Roberts and Tony Little (secondary supervisor, 2005 – 2006).

Craig Roberts and Thom Scott-Phillips deserve especial thanks for their insights, intelligence and inspiration.

All of the studies appearing in the thesis were devised in collaboration with Craig Roberts and carried out collaboratively with other researchers, as detailed below.

Part 1

The preamble to part 1 draws heavily on (Saxton et al. 2007), co-authored by Anna Lyndon, Anthony Little and Craig Roberts, and with especial thanks to Johan Lundström and Jan Havlíček for their helpful comments.

Chapter 2

Chapter 2 draws heavily on (Saxton et al. 2007; Saxton et al. 2008), both co-authored by Anna Lyndon, Anthony Little and Craig Roberts, with especial thanks Ajaz Iqbal and Fastlove Ltd for allowing use of participants from their speed-dating events, to Rebecca Fisher, Daniel Hughes, Chris Hassall and Alex Wall for their assistance with data collection, to Tom Heyes and Gregor Govan for laboratory support, to Thom Scott-Phillips and Johan Lundström for their useful discussion, to Anne Etgen and two anonymous reviewers for their helpful comments on the manuscript submitted to *Hormones and Behavior*, and to Unilever Research and Development for their generous funding of the project.

Chapter 3

Thanks to Beth Callinan and Penny Lewis for assistance with data collection and to Tom Heyes and Gregor Govan for laboratory support.

Chapter 4

Experiment 4a draws heavily on (Saxton et al. in review-a), co-authored by Rob Burriss, Alice Murray, Hannah Rowland and Craig Roberts, with especial thanks to Minna Lyons and Martin Guest for help with participant recruitment and session organisation, to Dave Perrett and Bert Tiddeman for use of the Psychomorph program, the Owen F. Aldis Scholarship Fund for its generous sponsorship of this research, and to anonymous reviewers for their helpful comments

Experiment 4b draws heavily on (Saxton et al. in review-c), co-authored by Ting Gao, Rob Burriss, Hannah Rowland, Jan Havlíček and Craig Roberts, with especial thanks to Minna Lyons and Martin Guest for help with participant recruitment and session organisation, to A. Kubena for advice on statistical analysis, to Dave Perrett and Bert Tiddeman for use of the Psychomorph program, and to anonymous reviewers for their helpful comments. This research was made possible by the generous award of the Owen F. Aldis Scholarship Fund to TS.

Chapter 5

Experiment 5a draws heavily on (Saxton et al. in prep.). Data collected in conjunction with Hannah Rowland and Ting Gao. Psychomorph used under licence from Dave Perrett. Tony Little created the facial images, provided the facial image examples in Figure 6 and wrote the first draft of the facial images methods section. Thanks to Kyrre Wathne and Steve Platek for use of their web platform (webexperiment.net) for online facial rating.

Chapters 6 and 7

These chapters draw heavily from (Saxton et al. in review-b), co-authored by Lisa DeBruine, Ben Jones, Tony Little and Craig Roberts. Data from Czech children were collected in collaboration with Dagmar Kohoutova and Jan Havlíček. Facial stimuli were created by Ben Jones. The authors thank David Feinberg (advice on voice stimuli creation), Thom Scott-Phillips (discussion and useful comments on earlier drafts), Andrew Boardman, Kristina Gilbertson, Gideon Gluckman, Shelly Kemp, Robert Kennedy, Jenny Saxton, Sue Toole (help with stimuli collection), the staff and pupils from George Heriot's School, the schools and social groups who provided much-valued participants, and the editor Martie Haselton and two anonymous reviewers for their helpful comments on the manuscript submitted to *Evolution and Human Behavior*. The images in Figure 8 were supplied by Lisa DeBruine and Ben Jones.

Experimental studies were approved by the University of Liverpool Committee on Research Ethics.

Chapter 1: An Introduction to Attraction and Attractiveness

> ***"Why does a particular maiden turn our wits so upside-down? The common man can only say, Of course we love the maiden, that beautiful soul clad in that perfect form, so palpably and flagrantly made for all eternity to be loved! And so, probably, does each animal feel about the particular things it tends to do in the presence of particular objects [..] To the broody hen the notion would probably seem monstrous that there should be a creature in the world to whom a nestful of eggs was not the utterly fascinating and precious and never-to-be-too-much-sat upon object which it is to her."***
>
> - William James (James 1890, p.261)

James' famous quote reminds us that physical attractiveness is not a property inherent to any individual, but rather is a sensation constructed independently in the mind of each viewer. When we label someone 'attractive', this is a short-hand way of saying that people tend to agree on how to rate their features. This point, that it is the perceiver who attributes attractiveness, is important, since it throws into focus two seemingly opposite but highly related questions: if attractiveness is rated independently, then how is it that people even of different ages, sexes, races, cultures and sexual orientations tend to agree on the relative attractiveness of others (Langlois et al. 2000)? And – on the basis that physical attraction plays a part in partner selection – why is it then that people *differ* in their choice of partner?

Animal mate choice can be described by general biological principles, shaped by the requirements of adaptive reproductive choice. Culture does not exempt humans from these principles: biological adaptation is still evident in their preferences and behaviour. The past few decades have seen a general trend for adaptive explanations of differences to be applied at increasingly fine-grained levels: from species, to local populations, to individuals (Wilson 1998). Differences between individuals, as much as differences between local populations or species, may be biologically adaptive (Qvarnström 2001). Differences within individuals may play a role in enhancing reproductive success, both in non-human animals and, most pertinently for this thesis, in humans.

This thesis examines some principal systematic sources of within-individual variation in human mate choice. If within-individual variation in choice is biologically meaningful and reflects selection for adaptive preferences, then we might expect to encounter its existence across a range of contexts and time scales. Accordingly, this thesis will consider how individual mate preferences vary in the short-, medium- and long-term time scales. It will report a range of experiments that consider how human mate choice changes according to cues available immediately in the environment, across the menstrual cycle, and during adolescence.

Before going on to do this, it is necessary to introduce the field. To begin with, in subsection 1.1 below, the general biological principles of mate choice are outlined, together with an explanation of why and how we encounter these principles across species. There will also be examples of how these principles play out in the specific case of humans. These examples are particularly important as they will reoccur frequently in the experiments reported in the thesis. Next, in subsection 1.2, we will see how individuals differ systematically in their choice of partners. This subsection will focus on between-individual differences. The empirical meat of the thesis will not concern itself with these types of differences, but they constitute an important methodological and theoretical foundation for the experiments carried out, and accordingly will be reviewed in some detail. Finally, in subsection 1.3, I introduce the topic central to this thesis: within-individual differences, and some of the ways in which they influence human mate choice. This variation will be divided up according to three different time scales: the short-, the medium- and the long-term changes that may have systematic effects on human mate choice. Against this background of variation in mate choice, subsection 1.4 explains how the thesis is structured, and the aims and objectives of the series of experiments undertaken.

1.1 Who is the best partner?

1.1.1 Attraction as adaptation

The sensation of attraction, cousin to sexual desire, enables humans to evaluate others in terms of how appropriate they are as a reproductive partner (Symons 1995). The world over, people assess attractiveness by evaluating indicators of the ability to provide the genetic or material resources necessary for healthy offspring (Symons 1995). In women, cues of high fertility, such as youth, the 'hourglass' figure, and signs of health, are universally desired; likewise, in men, signs of health, vigour, and resources are sought (Buss 1994). Since replication of one's own genetic material is crucially constrained by the selection of an appropriate partner, in so far as individuals are free to exercise any degree of choice in this matter, it is not surprising that the sensations which guide this choice bear the fingerprints of natural selection, in terms of heritable and complex design. The universality of human attractiveness judgments arises from the universality of human requirements in a mate.

Yet while we might all agree that the average model will tend to be more attractive than the average face on the street, there are also clearly between-individual differences in how we rate the desirability of potential partners. Undoubtedly, partner choice derives from a great deal more than physical traits. People tend to marry someone similar to them in terms of age, nationality, education and occupation (Jaffe & Chacon-Puignau 1995). Opportunity is also clearly essential; propinquity, or geographic proximity, is one of the greatest predictors of marriage partner. The immense importance of a shared history, and the value placed upon a suitable personality, character and other psychological traits, are significant influences on partner choice (Arum et al. 2008). Yet fascinatingly, all of this variance does not create sufficient noise to obscure the ways in which traits that are predominantly *physical* affect partner choice. Moreover, theorists link some of these individual differences in preference for physical traits to principles of optimal design. Differences in individual assessment regarding who is most attractive as a partner may arise not from individual miscalibrations of whatever internal mental apparatus assists in rating attractiveness, but may in some instances constitute an adaptive solution to an individual problem: not "Who is the best partner?", but rather, "Who is the best partner *for me*?"

Indeed, there are theoretical reasons why we might expect between- or within-individual differences to constitute individually optimal solutions. Evolutionary game theory (Maynard Smith 1979, 1982) can be used to formalise the advantages of various patterns or strategies of behaviour within a population. As neatly demonstrated by the 'hawk' and 'dove' division of strategies, for example, sometimes acting differently from your conspecifics can give rise to the richest pay-offs. Natural selection is sufficiently complex to instil sets of behaviour that vary from one individual to the next within a single species; strategic differences in reproductive behaviour, contingent upon differences in morphology, have been noted in a range of species. For instance, there exist many variations on the basic division between larger males who seek out females and compete with other males more aggressively, and smaller 'sneaker' males who mate opportunistically and covertly (review in Gross 1996). Pay-offs for different reproductive strategies may vary in relation to one's own phenotype, the local ecology and demography, one's status in relation to one's conspecifics, and the operational sex ratio; all of these factors have been demonstrated to affect reproductive behaviour in one species or another (Gross, 1996). The basic premise of strategic differences, that reproductive behaviour may be moulded by outside contingencies, has been applied to the relationship between humans and their mating behaviour.

1.1.2 Five abstract dimensions influencing mate choice

One main instigator of strategic differences in partnership preferences derives from the differences in reproductive investment between males and females (Trivers 1972). Men are able to limit their reproductive investment, at a bare minimum, to insemination, while the reproductive input of women requires lengthy gestation and nurture of the child. At the individual level, the consequences of reproductive choice can be serious. From the "gene's-eye" perspective (Williams 1966; Dawkins 1976), the consequences of choice of reproductive partner are even greater: they constitute the difference between continuation and annihilation. It is not surprising that the tools used in partner assessment, outlined next, are sophisticated and complex.

Overlying the basic differences between males and females is the fact that individuals differ in absolute terms in how desirable they are as a reproductive partner. Small differences in desirability can seem initially to be disproportionally related to reproductive consequences, particularly for males, whose mating success is usually more variable than that of females. At one level, variation is infinitely diverse: individuals may be highly desirable because of a fine set of tail feathers, a deep red skin coloration, or a solid jawline. Yet this variation has been reduced to a set of more abstract principles in an attempt to unite what is attractive across animal species. Making no especial claims to comprehensivity or orthogonality, five main tranches that capture a good deal of that variability are reviewed here. For each, we will see how these abstract principles apply across species; and then we will see specific examples of how they play out in humans. Many of the specific examples are introduced here because they will crop up over and again in the experiments reported later, and it is important to make a theoretical case for their usage. Determinants of physical attractiveness, then, may be captured with reference to fluctuating asymmetry, indicators of immunocompetence, averageness, health, and genetic heterozygosity.

First, symmetry is often attractive. In the absence of environmental perturbations, the instructions in the genome of an individual would give rise to a perfectly symmetrical phenotype (ignoring, of course, planned asymmetries, also known as directional asymmetry, such as the lobster's claw, designed to appear only on one side of the body). The amount of fluctuating asymmetry (FA) in a phenotype may act as an index of one aspect of genetic quality: the inherent ability of the genotype to withstand developmental upsets (see Lens et al. 2002 for an overview). Thus, mating with more symmetrical partners is likely to obtain heritable benefits for any offspring. Although the relationship between symmetry and reproductive quality has received much empirical support (Leung & Forbes 1996; Møller 1997), it is far from straightforward; for instance, studies which manipulated environmental stress to measure its effect on FA have shown that it may affect FA generally, in relation to specific traits only, or not at all (review in Bjorksten et al. 2000). While FA appears not to be the universal biometric anticipated by its earliest proponents, there still exists a wealth of evidence that it has some (varying) relationship with fitness across many taxa, providing useful information to conspecifics and researchers alike (review in Lens et al. 2002). In humans, symmetrical faces are preferred over asymmetric faces (e.g. Rhodes et al. 1998; Mealey et al. 1999; Perrett et al. 1999; Scheib et al. 1999; Little et al. 2001; Penton-Voak et al. 2001). This may not be due to symmetry *per se*: symmetrical faces are preferred even when raters see only half the face (Scheib et al. 1999). The authors suggest that this may be because measurable facial symmetry corresponds to other cues of attractiveness; these may include skin health (Penton-Voak et al. 2001; Jones et al. 2004b) or masculinity (Little et al. 2008 but see Penton-Voak et al.

2001). Further evidence for the correlation of symmetry with general attractiveness comes from findings that higher levels of symmetry correlate positively with greater vocal attractiveness (Hughes et al. 2002). Facial symmetry manipulations are used in the experiments described in Chapters 6 and 7.

A second dimension of desirability as a partner derives from markers of immunocompetence (Hamilton & Zuk 1982; Maynard Smith 1985; Folstad & Karter 1992). The immunocompetence handicap hypothesis applies the handicap principle (Zahavi 1975, 1977) to explain why markers of testosterone in males tend to be attractive (Folstad & Karter 1992). Testosterone suppresses the immune system (see e.g. Grossman 1984, 1985; Alexander & Stimson 1988; Folstad & Karter 1992; Rantala et al. 2003). They are thus a handicap to the bearer, and constitute an honest marker, or index, of genetic quality (Hamilton & Zuk 1982; Maynard Smith 1985; Folstad & Karter 1992), since individuals of lower quality cannot support their costs (Zahavi 1975, 1977). High testosterone levels do not go to fixation because of an ongoing 'arms' race' between parasites and their hosts (Dawkins & Krebs 1979). Despite a wide range of support for this theory, now accepted by many as canonical, it has not achieved universal acceptance; the occasional critics range from empiricists whose data do not support the assertion that testosterone suppresses the immune system (e.g. Klein et al. 1992) to theoreticians who point to the contradictory predictions of the immunocompetence handicap hypothesis: should high-testosterone individuals have stronger immune systems because they are able to overcome the handicap, or weaker immune systems but an ability to thrive nevertheless (Braude et al. 1999; Getty 2002)?

Despite these criticisms, markers of high levels of testosterone in males are often found attractive (Folstad & Karter 1992). In humans, testosterone levels covary positively with muscular development (e.g. Giorgi et al. 1999; Hansen et al. 1999), leading to more desirable body forms (e.g. Cohn & Adler 1992; Jackson & McGill 1996; Lorenzen et al. 2004). Likewise, women tend to like lower-pitched male voices (Collins 2000; Feinberg et al. 2005b; Feinberg et al. 2006; Saxton et al. 2006). Differences in voice pitch (or more strictly, fundamental frequency, of which pitch is the perceptual correlate) result from differences in the size and mass of the vocal chords (Titze 1994; Meredydd et al. 1998), and correspond negatively to differences in testosterone levels (Dabbs & Mallinger 1999; Evans et al. 2008) and positively to male reproductive success (Apicella et al. 2007). Women also prefer dominant behaviour (Dabbs 1997; Mazur & Booth 1998; Tremblay et al. 1998), again associated with higher levels of testosterone (Mazur & Booth 1998; Tremblay et al. 1998; Swaddle & Reierson 2002; Rowe et al. 2004; Archer 2006). The effects of facial markers of testosterone on preference present a slightly less clear picture. Women have been shown to prefer masculine faces (DeBruine et al. 2006), where markers of facial masculinity, including a more prominent brow and cheekbones and a larger chin (Grammer & Thornhill 1994), seem to develop under the influence of testosterone (Tanner 1978; Enlow 1990; Enlow & Hans 1996) and are associated with current levels of circulating testosterone (Penton-Voak & Chen 2004). However, other studies have shown a preference for a degree of femininity in male faces (Perrett et al. 1998; Rhodes et al. 2000; Penton-Voak et al. 2004; Jones et al. 2005a), perhaps because facial femininity is associated with traits that make for good partners or parents

(Perrett et al. 1998). Experimental procedures may influence these results to an extent; femininity preference appears stronger when the stimuli are created using facial morphing techniques rather than using real faces pre-rated for masculinity or femininity (Rhodes et al. 2000). However, there is a range of evidence to support the position that higher testosterone levels may mediate the allocation of mating effort or parenting effort (Archer 2006); men with higher testosterone tend to have a larger number of sexual partners, while lower testosterone is associated with the support of a partner within a monogamous relationship (Booth & Dabbs 1993; Gray et al. 2002; Burnham et al. 2003; McIntyre et al. 2006; van Anders et al. 2007; van Anders & Watson 2007; Peters et al. 2008).

It has been suggested that the handicap theory may also apply to markers of the "female hormone", oestrogen, in women (review in Rhodes et al. 2003a). Oestrogen is associated with the development of certain forms of cancer (Zeil & Finkle 1975; Colditz et al. 1995; Service 1998; Rodriguez et al. 2001), but is desirable in women: high levels of oestrogen correspond positively with increased reproductive potential (Stewart et al. 1993; Lipson & Ellison 1996; Baird et al. 1997; Baird et al. 1999) and increased female facial attractiveness and facial femininity (e.g. Law Smith et al. 2007). The handicap theory has been less well developed in females, perhaps because in most species they are the sex that chooses rather than the sex that displays (Darwin 1871). Manipulations of facial and vocal sexual dimorphism are used in the experiments reported in Chapters 5, 6 and 7.

A third dimension of desirability as a partner is derived with reference to averageness. Averageness is often associated with attractiveness, and may denote phenotypic efficiency: in a population, individuals who are more representative of that population are less likely to possess recessive genetic disorders and more likely to be better adapted to the demands of the local environment (Symons 1995). Averageness may also denote genetic heterozygosity, which has associated health benefits (Thornhill & Gangestad 1993), and is discussed in greater detail below. In humans, averageness has been most explored in relation to faces, where it tends to be found attractive (Langlois & Roggman 1990; Grammer & Thornhill 1994; Langlois et al. 1994; Rhodes et al. 1999; Rhodes et al. 2002; Rhodes et al. 2007; though see Grammer & Thornhill 1994; Pollard et al. 1999). Some recent studies also attempt to link vocal 'averageness' and attractiveness (Riding et al. 2006). The role of averageness in attraction is not wholly clear. The most attractive faces are not always the most average (Alley & Cunningham 1991; Perrett et al. 1994), and familiarity, which bears upon averageness, is also attractive outside of mate choice and in evolutionarily novel contexts (Halberstadt & Rhodes 2000; Rhodes et al. 2001; Rhodes et al. 2003b; Cooper et al. 2006; Quinn et al. 2008). Further, some aspect of the appeal of averageness may relate to ease of neurological processing (Winkielman et al. 2006), although of course this constitutes a proximate rather than ultimate explanation (Tinbergen 1963) of its appeal. Manipulations of facial averageness are used in the experiments reported in Chapters 6 and 7.

Fourthly, desirability as a partner may be influenced by health. Good health suggests both direct and indirect benefits: both that the individual is less likely to pass on infection at present; and also that heritable components of the individual's immune system may be more resistant to infection. In humans, ratings of perceived health covary positively with attractiveness (Fink et al. 2006) and FA (Shackelford & Larsen 1997). Attractiveness and actual health have also been linked directly when colour photographs (Shackelford & Larsen 1999) but not monochrome photographs (Shackelford & Larsen 1997; Kalick et al. 1998; Shackelford & Larsen 1999) are used. Although short-term variation in health is likely to affect attractiveness, better indices of general individual healthiness may be testosterone levels, FA or averageness (see e.g. Thornhill & Gangestad 1999a; Rhodes et al. 2003a), and these may be the more parsimonious predictors of attractiveness. Indeed, there is a great deal of covariance between a healthy appearance and the attractiveness of facial symmetry, sexual dimorphism and averageness (Rhodes et al. 2007; Scott et al. 2008).

Finally, links have been made between attractiveness and the major histocompatibility complex (MHC). The MHC, sometimes called the human leucocyte antigens (HLA) in humans, is the set of genes in vertebrates that code for immune response. MHC heterozygosity is controversially associated with health benefits (see Apanius et al. 1997 and Penn & Potts 1999 for conflicting viewpoints). Heterozygous individuals have been found to be more attractive relative to more homozygous individuals (Brown 1997, 1999). Diversity of the MHC may benefit disease resistance, by providing a 'moving target' for parasites, in a Red Queen-style arms race (Van Valen 1973); by enhancing heterozygosity and consequently the number of possibilities that the alleles available will be particularly well adapted to resisting local pathogens (Doherty & Zinkernagel 1975); or by increasing the diversity between parents, kin and offspring, such that individuals are less likely to be vulnerable to those parasites that are adapted to infest proximate individuals (see Thornhill et al. 2003). Alternatively, or perhaps in addition, MHC dissimilarity may simply act as an index of general genetic dissimilarity, and so selecting mates with dissimilar MHC might simply assist in the avoidance of matings with relatives and consequent inbreeding depression (Brown & Eklund 1994; Potts et al. 1994), although in humans other inbreeding-avoidance mechanisms such as taboo seem likely also to be influential. In humans, MHC heterozygosity is associated with higher ratings of facial attractiveness and apparent health (Roberts et al. 2005c), and with higher body odour ratings (Thornhill et al. 2003). Heterozygosity is used to investigate preference in the experiment reported in Chapter 3.

Each of these five abstract dimensions may be useful in that they can group together examples of attractive traits, while also explaining why such traits are attractive. Individuals who exhibit substantiations of these dimensions can in turn be subsumed by a more general term, that of high genetic quality, a phrase that encompasses 'good genes' (absolute quality) and 'compatible genes' (relative quality) (Neff & Pitcher 2005). This phrase is useful because of the hypothesis that attractive features advertise the genetic quality of the bearer (Thornhill & Gangestad 1999a), and indeed individuals who are attractive in one dimension tend to be attractive in others (review in Chapter 4).

Individuals differ in the extent to which they seek and value markers relating to at least some of these five dimensions of genetic quality in their partners (see subsection 1.1.2 below). There are other ways in which individuals differ, however, that, while desirable, are not universally so, since they relate to between-individual differences in genotype, phenotype or experience. Again, these between-individual differences in preference may fulfil a biological function. These are reviewed below in subsection 1.2. Although these sorts of differences are not the prime focus of the data chapters of the thesis, they present useful preliminary information, set the scene for some of the controls and possible sources of variance in the later experimental findings, and demonstrate again how general biological principles may apply in the case of humans. We have seen how mate quality may differ in absolute terms; we will now see how mate quality may differ in relative terms.

1.2 Who is the best partner *for me*? Variation between individuals

Systematic, between-individual differences in human mate choice again illustrate general biological principles. Among several important aspects, two factors affecting human mate choice are discussed here: assortative mating, and mate choice with reference to the MHC. These are chosen because they are biological factors that are better-established in terms of their effects, and also because they are relevant to the experimental designs in Chapters 3 and 5.

Some of the earliest empirical work on individual differences in mate choice judgments came about from the long-standing folk observation that people tend to resemble their partners. Empirical research shows that this type of positive assortative mating (also known as "homogamy") in humans exists both in psychological and in physical traits. Human partners tend to resemble each other in anthropometric, sociological and demographic measures such as age, height, weight, strength, attractiveness, visual similarity as rated by strangers, propensity to mental illness and ethnic background (for overviews and research see e.g. Spuhler 1968; Berscheid et al. 1971; Griffiths & Kunz 1973; Thiessen & Gregg 1980; Merikangas 1982; Malina et al. 1983; Zajonc et al. 1987; Hinsz 1989; Allison et al. 1996; Maes et al. 1998; Mathews & Reus 2001; Bereczkei et al. 2002; Bereczkei et al. 2004). Despite reports of physical similarity between partners, experimental work based on the experimental, computer-based manipulation of opposite-sex faces to increase self-resemblance has found a greater level of complexity in the ways that self-resemblance affects ratings of attractiveness. The earliest manipulation study found a positive relationship between self-resemblance and attractiveness, although it also found that the majority of this positive relationship could be explained by the positive effect of averageness on facial attractiveness (Penton-Voak et al. 1999b). A fairly large study of attractiveness ratings of male faces manipulated to resemble the rater showed that self-resemblance increased attributions of trustworthiness, had no effect on attractiveness for a long-term relationship, and actually decreased attractiveness for a short-term relationship (DeBruine 2005). However, preferences for self-similarity in both same- and opposite-sex faces have been demonstrated elsewhere (DeBruine 2004; DeBruine et al. 2005). Reports of physical similarity within real-life couples may represent a trade-off between attractiveness and trustworthiness in a partner (DeBruine 2005; DeBruine et al. 2008).

While social and cultural stratification undoubtedly accounts for some of the noted homogamy, there may also be more general biological forces at work; assortative mating has been demonstrated in a number of taxa (e.g. Cooke 1976; Davis & O'Donald 1976; Bateson 1978; Cooke 1978; Moravec et al. 2006). Research on assortative mating in humans has investigated where it is found, and also whether it is similar to the imprinting mechanisms demonstrated in sheep, goats, and birds (e.g. Oetting et al. 1995; Vos 1995; Kendrick et al. 1998), or to self-referent phenotype matching (seeking out mates who resemble one's self) as demonstrated in peacocks (Petrie et al. 1999).

Investigation of visual trait preference in humans suggests there is a stronger component of something akin to 'imprinting', and a lesser component of self-referent phenotype matching. Two studies of real-life marriage choices have shown greater similarity between a target's spouse and his or her opposite-sex parent than between the target and his or her spouse (Bereczkei et al. 2002; Bereczkei et al. 2004). The second of these studies used adoptive families to control for target-parent resemblance, although this does not entirely solve the problem; people report greater willingness to favour or adopt children who resemble them (Platek et al. 2002; Bressan et al. 2008). There was also evidence for a small degree of physical similarity between the adopted daughter and her husband, i.e., 'own-phenotype' matching (Bereczkei et al. 2002). Another study on visual characteristics, this time on hair and eye colour, found that opposite-sex parental characteristics were most important in predicting partner characteristics (i.e. 'imprinting' effects); there was also evidence for some self-referent phenotype matching for hair and eye colour (i.e. 'own-phenotype' matching) (Little et al. 2003). In the olfactory modality, one study found that women preferred body odours associated with the set of immune-system encoding genes inherited from their fathers but not their mothers (Jacob et al. 2002b), although this study was a little limited by its small sample size.

The term "imprinting" may not be the best descriptor of the process; we should perhaps refer to these effects in humans as forms of *social learning,* or *familiarisation* (McFarland 1993) since the term imprinting usually denotes a bout of fixed and irreversible learning within a limited sensitive period, which does not seem to be the case in humans (Little et al. 2003). For instance, earlier findings that parental age may have a very small positive effect on the partner a female actually chooses (Zei et al. 1981; Wilson & Barrett 1987) have been followed up with an investigation of preference (Perrett et al. 2002). Older age was less aversive to people with older parents than to people with younger parents. As the authors noted, this tolerance to cues of age cannot be derived directly from a period of infanthood exposure to older parental faces, since the parents would have been younger than the rated faces during that period, and in this way these effects in humans differ from the standard definition of imprinting.

The finding that real-life partners exhibit physical similarity (Spuhler 1968; Zajonc et al. 1987; Bereczkei et al. 2002; Bereczkei et al. 2004) may be because cues of a degree of relatedness are beneficial in a long-term relationship. A preference for some degree of similarity in one's mate may enable optimal outbreeding (Bateson 1978, 1980, 1982), encourage the selection of a partner from the same population who is more likely to have appropriate adaptations to the local environment thereby enabling the maintenance of co-adapted genetic complexes (Read & Harvey 1991), and enhance one's own genetic representation in future generations through the selection of a partner with some genetic matches (Thiessen 1999). Long-term relationships may benefit from a degree of affiliative behaviour associated with distant relatedness, including enhanced trust (DeBruine 2005). Together, these things may outweigh the aversion of a small degree of self-resemblance (DeBruine 2005; DeBruine et al. 2008). This question is take up again in Chapter 5.

The second biological principle to be discussed here relates to the MHC. The MHC was introduced above; in general, MHC heterozygosity is attractive. Yet the MHC may also influence preference at a relative level. There is some evidence that humans, like many other vertebrates, mate disassortatively with regards to their MHC type (Penn 2002; Bernatchez & Landry 2003), a practice that may function to reduce inbreeding depression, where the viability and health of the offspring is reduced (Potts & Wakeland 1993). Although MHC-disassortative mating in humans remains to some extent contested (e.g. Nordlander et al. 1983; Hedrick & Black 1997), evidence in its favour arises from studies of both preferences and behaviours. The mechanism by which conspecifics judge MHC dissimilarity may derive primarily from body odour (Penn & Potts 1998): mice, rats and humans can all distinguish very slight differences in MHC by odour cues (Yamazaki et al. 1983; Penn & Potts 1998; Jacob et al. 2002b). Thus, odour preference has been used to measure human MHC preferences.

Odour preference studies in men and women have tended to show negative correlations between rated pleasantness of body odour and extent of MHC similarity between rater and donor (Wedekind et al. 1995; Wedekind & Furi 1997; Thornhill et al. 2003), although an intermediate level of dissimilarity may be more desirable than extreme dissimilarity (Jacob et al. 2002b), perhaps because excessive diversity of the MHC may reduce immune functioning (Nowak et al. 1992). Behavioural studies have also lent some tentative support to MHC-disassortative mating. Couples within genetically non-diverse populations (i.e. chosen to reduce the range of MHC types and attendant confounds and methodological difficulties) are more dissimilar than would be expected by chance (Ober et al. 1997). Likewise, children within an isolated population were less homozygous at the MHC than would be expected, implicating selectional abortion or difficulty conceiving among MHC-similar couples (Kostyu et al. 1993; Ober 1995), and pregnancies of MHC-similar couples are associated with higher levels of spontaneous abortions (Komlos et al. 1977; Schachter et al. 1984; Jin et al. 1995; Ober et al. 1998). However, other studies have presented conflicting results; for instance, one study showed a greater number of MHC-*similar* partnerships in a study of a large number of couples than would be expected by chance, although this is possibly due to ethnic and racial positive assortative mating within the diverse population sample (Rosenberg et al. 1983).

The only work on facial preferences that took account of the MHC of the rater and the rated face found that females preferred the faces of males whose MHC was more, rather than less, similar (Roberts et al. 2005b). The authors suggested that MHC-related information from the different channels (vision and odour) could filter out potential mates of very similar and very dissimilar genotypes respectively. Certainly, in so far as odour plays a role in mate choice, it must be an interactive rather than primary role, as demonstrated by the consistent inability of individuals even to judge sex accurately by odour (Wedekind et al. 1995; Wedekind & Furi 1997; but see Russell 1976).

These between-individual differences demonstrate how reproductive theory from traditional biology is relevant to human partner preferences. Individuals may evaluate mate quality in relative terms, and these differences may be biologically adaptive. An awareness of systematic between-individual differences in preference is essential to methodological considerations in empirical studies of mate preferences. The empirical focus of the thesis, though, constitutes within-individual differences, and we turn to those now.

1.3 Who is the best partner *for me*? Variation within individuals

We have seen some examples of sources of between-individual differences in variation for partnership preference. Yet people also vary in their preferences over time. These differences are in some senses more complex, relying on an iterative process between the individual and the natural and social environment. Again, these individual differences may help optimise reproductive choice (Via & Lande 1985); in a changing environment, the ability to modify behaviour in response to outside cues may be particularly useful (see e.g. Wilson et al. 1994; Qvarnström 2001). If there is a general biological principle allowing for facultative adjustment of human mate choice decisions, and of course subject to biological constraints on plasticity (Via & Lande 1985; DeWitt et al. 1998), we might expect to see this at work across short-, medium- and long-term time scales. Existing evidence for such adjustments is discussed here. Adjustments may be along any of the five dimensions outlined in subsection 1.1.2, but in fact most work on within-individual preference differences has been carried out in relation to preference for symmetry and testosterone markers, and these are the focus here.

1.3.1 Within-individual changes across time: short term time scales

Different contexts provoke differences in preference; context shifts thus can lead to immediate preference shifts. This has been demonstrated extensively in relation to the contexts of short-term and long-term sexual relationships. (The standard terminology contrasts short-term and long-term relationships; this thesis structure contrasts within-individual changes at long-, medium- and short-term time spans. The inconvenience of using 'short-term' in these two different senses is outweighed by the advantage of maintaining the standard terminology). Long-term sexual relationships are a central feature of the societal make-up of every known culture, involving often extensive marriage traditions. Yet short-term relationships are also a characteristic of human society. Individuals can take part in dates, blind dates, speed-dating and holiday romances. Outside of western societies, the universal prevalence of cultural or religious prohibitions to short-term relationships, such as adultery, attest to their very existence. As such, the within-individual changes in preference that can be revealed by asking the individual to make judgments pertaining to a short-term and then a long-term relationship context constitute a good demonstration of preference changes over the short term time span.

When making judgments for short-term compared with long-term relationships, a number of studies have demonstrated that women show a greater preference for lower levels of male facial femininity or higher levels of facial masculinity (Penton-Voak et al. 1999a;

Little et al. 2002; Penton-Voak et al. 2003). In short-term but not long-term relationships, women also prefer lower-pitched voices (Puts 2005), and more masculine behaviour (Gangestad et al. 2004). Convincingly, together, these studies demonstrate a preference for testosterone markers for short-term compared with long-term relationships over a range of modalities. In terms of men's preferences for women, the most attractive female waist-to-hip ratio is also more salient when judgments are made for a short-term relationship (Schmalt 2006).

While partners in long-term relationships are expected to exhibit high levels of commitment on both sides, short-term relationships are usually of a more sexual nature. When individuals consider sexual relationships, they may privilege indirect benefits (genetic quality) over direct benefits (resource provision and commitment) (Buss & Schmitt 1993; Buss 1994; Gangestad & Simpson 2000; Li & Kenrick 2006). This is a particularly apposite distinction since there is evidence for an inverse correlation between male provision of direct and indirect benefits. For instance, more symmetric men report lesser willingness to provide gifts (Gangestad & Simpson 2000). Similarly, partnered men tend to have lower levels of testosterone than single men, and among partnered men, lower testosterone correlates with higher investment in the relationship and higher levels of marital stability (Gray et al. 2002; Gray et al. 2004). The differences between preferences for short- and long-term relationships have been worked into the strategic pluralism hypothesis (Gangestad & Simpson 2000). A critique of the controversies of this hypothesis (see e.g. Cornwell et al. 2002) is beyond the scope of this review, since here we focus on preference, not behaviour. It is enough to say that irrespective of the functions of the differences in preference, asking individuals to focus alternatively on short-term and long-term relationships does seem to give rise to systematic differences in preference, as we see again in Chapters 3 and 5.

Another way in which individual preferences can change over a short-term time scale is in response to sensory modulation within the immediate environment. For instance, hungry raters demonstrate an increased preference for figures of women with a greater body weight (Swami & Tovée 2006). Likewise, the presentation of a series of faces that have been exaggerated along one dimension leads to increased ratings of normality for similar-looking faces (Little et al. 2005). Odour cues may also affect face ratings over short-term time spans (Li et al. 2007), and it is this topic that is explored in depth in Chapters 2 and 3.

1.3.2 Within-individual changes across time: medium-term time scales

Individuals may vary in their preferences across medium-term time scales, that is, in respect of changes that are likely to occur a number of times across the life course. Relationship status and self-perceived mate quality, discussed below, are chosen here as examples of how preferences may change over medium-term time scales, because these have been investigated by a range of studies, because they have been interpreted according to general biological principles relating to the enhancement of reproductive success, and because these differences will be referenced in the data chapters that follow.

Relationship status has been found to influence preference. Facial masculinity is preferred more by women in a relationship than by unpartnered women (Little et al. 2002). Similarly, while women tend to prefer greater levels of facial masculinity at the more fertile phase of the cycle (see Chapter 5), women within relationships show a more pronounced cross-menstrual cycle preference shift away from faces that indicate lower levels of testosterone (Penton-Voak et al. 1999a; Little et al. 2002). These differences between partnered and non-partnered women have been used to support the strategic pluralism hypothesis (Gangestad & Simpson 2000): that is, an increased awareness of mate quality as demonstrated by the women in relationships may allow them to obtain indirect benefits from conceptive extra-pair copulations (Gangestad & Simpson 2000; Little et al. 2002). It is worth noting that in the absence of within-subjects studies on relationship context, we do not know whether these differences arose after the relationship began (as presumed), or whether there are systematic differences between preference and the likelihood that a woman is in a relationship.

Own mate quality also predicts preferences. One study (Pawlowski & Dunbar) showed that human mating preferences can be described in terms of a marketplace, where more desirable individuals can 'afford' more desirable partners. The study used 'lonely hearts' advertisements, and so was somewhat restricted in its assessment of individual demands and desirability by the limitations of that format. Nevertheless, it was able to conclude that women's 'market value' could be described in terms of fecundity, and men's 'market value' in terms of earning potential and the risk of future pairbond termination (by death or divorce). These conclusions chime with other predictions of reproductive theory: fecundity constitutes more of a reproductive limit for females, who have a shorter reproductive life than men and a long gestation period. In contrast, men are limited by female access; male resources and commitment may enhance access to females. Corroborative evidence for the existence of partnership stratification by quality comes from female evaluations of past partners, where the resource potential, commitment and desirability of personality of a woman's partner have been shown to be positively related to that woman's attractiveness (Mathes & Kozak 2008). Likewise, choosiness for attributes previously demonstrated to be important to mate choice (e.g. wealth, status) was found to be positively correlated with higher ratings of one's own possession of those attributes, in both males and females (Buston & Emlen 2003), and thus possession of overly high socio-economic status may be an aversive attribute in a long-term relationship for many (Chu et al. 2007). The positive relationship between mate desirability and choosiness has also been captured in a mathematical model (Hill & Reeve 2004).

Mate quality in turn affects preference for specific physical traits. Women's choosiness (i.e. increased preference for facial masculinity and symmetry) has been shown to vary positively with self-rated attractiveness when making judgments of male but not female faces, and for a long-term but not short-term relationship (Little et al. 2001). Since the original study, this pattern has been replicated on some occasions (Little et al. 2007b) but not others (see e.g. Penton-Voak et al. 2003; Jones et al. 2005b; Cornwell et al. 2006; Little et al. 2007a); it is not clear whether this is because the relationship between self-rated attractiveness and choosiness is complex, or because of a confounding link such as that between self-rated attractiveness and menstrual cycle stage (e.g. Penton-Voak et al. 2003; Schwarz & Hassebrauck 2008).

Nevertheless, the relationship between some measure of own mate quality and choosiness has been replicated in other ways. Women who were judged by others to be more attractive or to have a more attractive body shape showed a stronger preference for less feminised faces in the context of a long-term relationship (Penton-Voak et al. 2003). Women with more attractive waist-to-hip ratios and body weight showed stronger preferences for male resources, attractiveness and health (Jones et al. 2005b; Pawlowski & Jasienska 2008). Likewise, women with greater psychological health (i.e. lower scores on measures of anxiety, stress and depression), showed greater attraction to cues of health in male faces (Jones et al. 2005b). Greater health in both men and women has been shown to predict preferences for facial sexual dimorphism in a rural population (Scott et al. 2008), and men's testosterone levels correlate positively with their preference for femininity in female faces (Welling et al. in press). A contrary finding, that women with a less attractive body weight were more demanding in their preferences for male attractiveness (Pawlowski & Jasienska 2008), may be due to some confound with a specific cultural stratification that links BMI and culturally-induced choosiness. In both males and females, age of first sexual encounter is associated positively with greater preference for mate quality, this time as measured by facial sexually dimorphic characters and facial attractiveness (Cornwell et al. 2006). Further, cueing an individual to believe that the social environment contains a large number of high-quality same-sex others can lead to decreased choosiness in mate choice (Little & Mannion 2006).

There are a number of non-exclusive explanations for preference variation associated with mate quality. At the proximate level, self-esteem may link one's own attractiveness to the pursual of more attractive partners. If all agents in a population seek out the best partner available, even without pre-existing knowledge of one's own quality, the individuals making up a partnership will be similar in quality (see e.g. Johnstone 1997). Preference adjustment according to own desirability may be adaptive: once own mate value has been learnt, time is saved by seeking partners amongst those who are somewhat likely to acquiesce. Such stratification by quality on the basis of a heuristic along the lines, 'Obtain the best quality possible in a reasonable amount of time' may not be restricted to reproductive choices; we should find the same sort of stratification in any population where each individual agent aims to obtain the best quality possible. Condition-dependent mate preferences have been documented in a range of species. For instance, female zebra finches sought more or less attractive mates depending on their own quality, which was experimentally manipulated through the clipping of wing feathers (Burley & Foster 2006). Likewise, female fish in poor condition exhibited a lower level of choosiness with reference to male coloration, a marker of quality, in both guppy fish (*Poecilia reticulata)* (Lopez 1999) and three-spined sticklebacks (*Gasterosteus aculeatus*) (Bakker et al. 1999). Female guppy fish condition also correlates with level of preference for high-quality males with more symmetric spines (Mazzi et al. 2003).

In humans, hypotheses linking reproductive quality and adaptive preferences have been formulated, drawing on the findings of condition-dependent preferences in long-term but not short-term relationships, and in respect of male but not female stimuli. It has been argued that lower-quality women may obtain more from direct benefits such as stable investment (from less reproductively desirable men) than from indirect benefits, or alternatively that only higher-quality females have sufficient resources themselves to be able to risk male desertion, something which is less likely with less attractive males (Little et al. 2001; Little et al. 2002; Penton-Voak et al. 2003; Feinberg et al. 2006). More complex relationships between a woman's mate quality and preference variations across the menstrual cycle have also been demonstrated; these are discussed in Chapter 5.

1.3.3 Within-individual changes across time: long-term time scales

This section reviews systematic within-individual differences that may occur over a longer time scale. This category discusses the links between preference and socio-sexual orientation, environmental clemency, and age.

The socio-sexual orientation index, or SOI, sets out to measure individual openness towards uncommitted sexual relations. Individual SOI may be described as a long-term strategy that develops with reference to social and environmental context (Schmitt 2005). Women with higher SOI are more likely to prefer masculine faces, where masculinity was determined by objective measurements of facial trait sizes (Waynforth et al. 2005), and males with higher SOI are more likely to prefer the most attractive female waist-to-hip ratio (Brase & Walker 2004). These findings may also reflect the documented greater preferences for mate quality for short-term relationships.

Environmental harshness, which has previously been shown to affect general-level reproductive strategy (Belsky et al. 1991; Schmitt 2005; Belsky et al. 2007), also affects preference for partner characteristics. A vignette-based study showed that participants who imagined themselves in a safe environment showed stronger preferences for greater levels of opposite-sex facial sexual dimorphism (masculinity in male faces or femininity in female faces) in long-term but not short-term relationships compared with participants who imagined themselves in a harsh environment (Little et al. 2007a). In harsh environments, it may be more important that the partner should invest heavily in the relationship and any offspring, which may give rise to the preference for lowered facial sexual dimorphism (Little et al. 2007a). Consistent with this, cultures with harsher environmental conditions seem more likely to be those cultures with greater expectation of higher levels of commitment prior to sexual relationships (Gangestad & Simpson 2000; Schmitt 2005). However, increased parasite load, which may also be associated with a harsher environment, is associated with increased preferences for greater levels of sexual dimorphism in faces (Penton-Voak et al. 2004) and a greater emphasis on physical attractiveness (Gangestad & Buss 1993); both preferences may function to increase attraction to healthier partners (Gangestad & Buss 1993; Penton-Voak et al. 2004).

Finally, age may provoke changes in partnership preferences across longer time spans. Most research has tended to use university undergraduates as experimental participants, and age-related preference changes have been little examined. The long-term mate choice changes that take place during adolescence will be a major focus of this thesis, in Section 3.

1.4 Thesis outline and objectives

We have seen how preference differences may relate to absolute or to relative mate quality. Preferences may vary between- or within-individuals. Within-individual differences may occur on a short-term basis, in response to cues immediately available in the environment; over a medium-term time scale, with changes which may occur regularly; and across a longer period of time. Each of these time scales will form the stage for empirical investigations of within-individual variation in partner choice within this thesis, with the overall aim of showing that such time scales provide a useful framework to examine individual variation in human mate choice.

In the short term, mate choice may be adjusted facultatively with reference to specific odour cues (Section 1). Chapter 2 examines the effects of androstadienone, a chemical constituent of human sweat that may be secreted in higher concentrations by males than females (see Gower & Ruparelia 1993 for an overview). Androstadienone has been linked to a number of psychological and physiological changes, particularly in females (review in Saxton et al. 2007). Yet it has been little examined in ecologically valid contexts, a flaw for any who would put it forward as a potential human chemosignal or pheromone (e.g. Grosser et al. 2000; Jacob & McClintock 2000; Jacob et al. 2001a; Jacob et al. 2001b; Lundström et al. 2003a; Gulyas et al. 2004; Lundström et al. 2006). Chapter 2 reports the results from an investigation into whether this chemical can affect women's perceptions of men's attractiveness at a speed-dating event, an environment chosen for its ecological validity. It finds that women exposed to androstadienone give higher ratings of the attractiveness of the men they interact with, compared to women in both odour and non-odour control conditions.

Chapter 3 takes up two of the questions raised by these findings. Androstadienone is perceived in the olfactory modality, broadly speaking, and Chapter 3 firstly reports an experiment investigating whether female ratings of male attractiveness can be modulated in the olfactory modality as much as in the visual modality. It again finds that androstadienone has a positive effect on ratings, with some evidence of a stronger effect in the visual channel. If androstadienone can affect perceptions of attractiveness, one obvious question is whether change applies to all men, or to men of greater or lesser mate quality. This is examined by asking women to rate men who have been genotyped for MHC heterozygosity. In this experiment, there is again evidence that androstadienone enhances ratings of attractiveness. There is some evidence for an interaction between the genetic quality of the stimulus and the extent of attractiveness enhancement, although this is inconsistent. The experiment also raises methodological issues regarding the standard concentration of androstadienone used in tests of its effects. With regard to these results, the general discussion to Section 1 emphasises the importance of determining whether androstadienone production rates in men can be shown empirically to provide information relating to mate quality.

Section 2 addresses within-individual variation over the medium-term, with a particular focus on the menstrual cycle. Despite a large number of studies on how menstrual cycle phase relates to attractiveness of appearance, none has tried to quantify the relative contributions of the different channels in terms of attractiveness changes. Chapter 4 firstly reports an experiment to quantify the relative importance of the different modalities in rating attractiveness, irrespective of cycle phase. It finds that both visual and vocal attractiveness are good predictors of overall attractiveness, thereby recommending an increased focus on vocal attractiveness to the research community. The second experiment goes on to attempt to quantify the relative importance of the different channels in terms of the documented increases in attractiveness associated with the menstrual cycle. It finds that attractiveness changes in the predicted direction are detectable in a range of modalities, with some evidence that they may be clearest in facial appearance.

Menstrual cycle phase has also been linked to changes in mate-related preferences, especially for self-resemblance and masculinity (review in Jones et al. 2008a). Self-resemblance and masculinity represent cues of relative and absolute mate quality respectively (Roberts & Little 2008). In a novel design recommended in (Roberts & Little 2008), both dimensions are manipulated simultaneously in facial photographs to determine how they affect female preference, and how this may change with the menstrual cycle. Evidence is put forward for the presence of a context-dependent heirarchical rule trading off genetic quality and masculinity.

Finally, adolescence is used to investigate within-individual preference changes across the long-term time span (Section 3). Adolescence is a particularly interesting time from the point of view of biological adaptation since it constitutes the period when an individual becomes reproductively capable. Chapter 6 asks how biological development, captured by chronological age and pubertal development, predicts preferences for faces and voices. It finds that adult-like preferences develop across age, and also that preference for facial cues of mate quality are evident prior to the reproductive years. It also gives some evidence of individual differences in attractiveness judgments corresponding to individual differences in pubertal development, controlling for age.

Chapter 7 then looks at the moderating effects of experience on preferences over adolescence. It examines whether the environment (here, mixed- contrasted with single-sex school, and presence or absence of siblings) can systematically modify preferences. It finds some evidence for experience-based effects on preferences.

Finally, Chapter 8 recaps these findings, and discusses their wider implications for the field of evolutionary psychology. Criticisms have been levelled at evolutionary psychology as a research paradigm, yet this paradigm has enabled the productive investigation, in the foregoing data chapters, of individual variation in human mate choice. Indeed, many of the criticisms of evolutionary psychology can in fact fission productive sources of research. In the distant future, work such as that set out in these data chapters may contribute, albeit minutely, to any epistemological assessment of the evolutionary psychology research paradigm.

SECTION 1: SHORT-TERM VARIATION: FACULTATIVE ADJUSTMENT TO ODOUR STIMULI

Part 1 reports three experiments that investigate how human mating preferences may be modulated over the short term, in response to an odorous chemical. Some of the indirect evidence that odour is used in human partnership choice and formation was reviewed in Chapter 1. Women in particular report odour to be important in partnership choice and maintenance (Herz & Cahill 1997; Herz & Inzlicht 2002; Havlicek et al. 2008), and MHC-linked odours appear to affect human mate preference and choice (reviews in e.g. Potts & Wakeland 1993; Penn & Potts 1998; Penn & Potts 1999; Penn 2002). Further, odorous chemicals are used extensively in mate choice in non-human animals, and constitute a rich and potent source of potential information to conspecifics. This information can include sex, identity, relatedness and individual condition, and can be gleaned from glandular sources and urine or fecal scent marks (Gosling & Roberts 2001). In general, exposure to odours can lead to immediate modulation of perception, cognition and behaviour (reviews in e.g. Chu & Downes 2000; Holland et al. 2005). Together, these properties make odour a good contender for an agent to investigate human preference changes over the short term.

One approach to understanding how humans are influenced by biologically relevant odours has focussed on the information accessible through chemosignals, without specific knowledge about the chemical substances involved. For instance, this approach has investigated the relationships between hedonicity of axillary odour and partner preferences based on genetic similarity

(Wedekind et al. 1995), the effects of chemical secretions on the timing of the female menstrual cycle (McClintock 1971, 1998), and the relationship between personality and odour (Havlicek et al. 2005). An alternate approach has quantified the action of specific chemicals of biological origin on behavioural or psychological states. Chemicals in this class have sometimes been referred to as "chemosignals" or "pheromones".

Pheromones have been defined as substances released by an individual into the external environment which precipitate a particular reaction in a conspecific (Karlson & Lüscher 1959). Pheromones are used by species in a variety of phyla (see e.g. McClintock et al. 2001), and there exist many examples of pheromone-mediated behaviour in a wide range of mammals, particularly in relation to mating behaviour and maturation (see e.g. Vandenbergh 1983b, a). In humans however, the question of whether pheromones influence behaviour was recently listed by Science magazine as one of the top 100 outstanding scientific questions (Anon. 2005). A recent review of behavioural and anatomical studies relating to the function of pheromones in human interactions concluded that while a small number were "unambiguously supportive", none seemed ultimately conclusive (Hays 2003). In non-human animals, pheromones are associated with direct and specific behavioural effects (Karlson & Lüscher 1959), allowing them to be sub-classified as types such as primers or releasers. The absence of such direct behavioural changes in humans has led to the suggestion that purported human pheromones may be better conceived of as modulatory chemosignals (Jacob & McClintock 2000; McClintock 2000).

One hurdle for proponents of human pheromones is the lack of clear evidence for a functional human vomeronasal organ (VNO). Located within the nasal cavity, the VNO is used by many terrestrial vertebrates to detect pheromones (Johns et al. 1978; Keverne 1983). Although one study (Grosser et al. 2000) reported measurable autonomic changes following direct introduction of a proposed human pheromone to the human VNO such that this chemical was prevented from reaching the main olfactory epithelium, the VNO system appears to be vestigial in humans. The VNO is not consistently identified in human adults (Garcia-Velasco & Mondragon 1991; Moran et al. 1991; Stensaas et al. 1991; Trotier et al. 2000; Bhatnagar & Smith 2001; Knecht et al. 2001); a discernable physiological substrate from the VNO to the brain has not yet been identified (review in e.g. Meredith 2001); and those VNO receptor genes which we might expect to enable pheromone detection (on the basis of comparative studies) appear to be non-functional or pseudo-genes in humans (Tirindelli et al. 1998; Kouro-Mehr et al. 2001). However, pheromonal reception may be possible through the main olfactory system (see e.g. Dorries et al. 1997; Restrepo et al. 2004; Liberles & Buck 2006), and so continuing debate surrounding the human VNO should not preclude ongoing research into the behavioural effects of potential human pheromones (Meredith 2001).

To date, work on possible human pheromones has primarily focused on the androgens androstenone, androstadienone, and androstenol, found in the human axilla (review in Hays 2003). These may be found in higher quantities in males than females (see Gower & Ruparelia 1993 for an overview). Androstenol production rates seem to rise at puberty (Cleveland & Savard 1964) and fall in old age and post-menopausally (Brooksbank &

Haslewood 1961), in a pattern of ontogeny and sexual dimorphism consistent with sexual signalling. Metabolites of androgens, 5α-androstenone, androstadienone and the low-odour androstadienol, are secreted from the apocrine glands (Gower et al. 1994). Androstadienol and androstadienone have been shown, *in vitro*, to be converted to the musky-smelling androstenol and the more prominent, urinous androstenone, under the influence of skin-inhabiting coryneform bacteria (Mallet et al. 1991; Gower et al. 1994), although alternative biotransformational pathways have also been demonstrated, again *in vitro* (Decreau et al. 2003).

The earliest work on "human pheromones" was mostly concerned with androstenone and androstenol, perhaps because of their documented function in other mammals: androstenone, for example, occurs in boar saliva and triggers lordosis (the 'mating stance') in the female pig (Signoret & du Mesnil du Buisson 1961). Yet the results of these earlier experiments do not lead to a consistent picture of the effects of these two chemicals in humans. For example, while some studies purport to show that androstenone or androstenol exposure reduces female ratings of male sexual attractiveness (Filsinger et al. 1985), others show that androstenol increases female ratings of character or sexual attractiveness (Cowley et al. 1977; Kirk-Smith et al. 1978), or that there is no effect of either chemical on judgments (Black & Biron 1982; Filsinger et al. 1984). In terms of emotional and behavioural responses, female exposure to androstenone has been associated with a decrease in her self-perceived "sexiness" (Filsinger et al. 1984), but with an increase in her levels of social interactions with males (Cowley & Brooksbank 1991). On the other hand, neither androstenol nor androstenone led to increased female sexual arousal in response to erotic prose (McCollough et al. 1981;

Benton & Wastell 1986). Androstenol exposure has been linked with increased mid-cycle self-reported feelings of submissiveness (Benton 1982) and irritability during menses (Cowley et al. 1980), and indeed its effects on the menstrual cycle may play a part in the reported menstrual synchrony of women who live together (Shinohara et al. 2000). In two studies, females chose to approach areas which had been dosed with androstenone (Kirk-Smith & Booth 1980; Pause 2004); in a similar study, androstenol had no such effect (Gustavson et al. 1987). The effects of these chemicals on males are sometimes, but not always, complementary: in Pause's (2004) study, homosexual men, like women, approached areas dosed with androstenone more often (heterosexual men were not tested). In another study, men (whose sexual orientation was not reported) rated other men as more sexually attractive in the presence of androstenone and androstenol (Filsinger et al. 1985). Although few of these studies on pheromones in humans are directly contradictory, they are sufficiently at odds to indicate that we lack enough understanding of the phenomenon to be able to ask coherent questions of its effects.

One review of these findings notes that only some of these studies have used a musk odour control: it could be that all musk odours would trigger similar responses to these chemicals (Gower et al. 1988). Nevertheless, an attenuated conclusion from the work would merely be that some musks have behavioural effects, and androstenol and androstenone are examples of such musks. An alternative suggestion to help explain the reported inconsistencies is that these chemicals have different effects depending on the contexts in which they are experienced, and furthermore that only within more ecologically valid conditions are we most likely to detect their influences (Jacob et al. 2002a; Saxton et al. 2007).

This important issue will be taken up again in Chapter 2 with a study of a purported human pheromone that has received the most research interest recently: androstadienone.

Chapter 2 reports the results from three studies of the effects of androstadienone at speed-dating events in which men and women interacted in a series of brief dyadic encounters. Men were rated more attractive when assessed by women who had been exposed to androstadienone, although this result was apparent in only two out of three studies. Selection rates were not affected by androstadienone exposure. Chapter 3 reports a study where the effects of androstadienone on attractiveness ratings in the olfactory and visual modalities were contrasted. In the first experiment, women rated the attractiveness of male body odour samples, facial photographs, and odour/photograph combinations in the presence of androstadienone or a control in a within-subjects experiment. There was a significant increase in ratings in the androstadienone condition. In the second experiment, women rated a set of facial photographs of males of known MHC heterozygosity, a measure of genetic quality, in the presence of androstadienone or a control in a within-subjects design. There was some evidence for the enhancement of male attractiveness when assessed in facial photographs. Interactions with underlying male attractiveness or MHC type were also investigated.

Chapter 2: Androstadienone effects outside the laboratory

This chapter describes an experiment designed to test the effects of androstadienone, a component of human sweat, on women's perceptions of men's attractiveness. Androstadienone has been associated with a number of moderately consistent effects. Women exposed to androstadienone in the laboratory environment tend to experience positive affective modulation, including fewer feelings of boredom and frustration associated with the laboratory testing session, maintenance of positive moods despite exposure to negative stimuli, and increased feelings of focus (Grosser et al. 2000; Jacob & McClintock 2000; Jacob et al. 2001a; Lundström et al. 2003a; Bensafi et al. 2004a; Lundström & Olsson 2005; Villemure & Bushnell 2007). At the physiological level, androstadienone inhalation precedes measurable changes in endocrine state (Wyart et al. 2007) and autonomic activations (Monti-Bloch & Grosser 1991; Grosser et al. 2000; Jacob et al. 2001a; Bensafi et al. 2004a) that may be specific to females (Boulkroune et al. 2007) or to individuals who are sexually attracted to men (Savic et al. 2005). Neurologically, female response to androstadienone extends beyond the olfactory system, activating areas of the brain associated with attention, social cognition, emotional processing, and sexual behaviour (Jacob et al. 2001b; Savic et al. 2001; Gulyas et al. 2004; Savic et al. 2005). A functional consideration of these types of findings has led to the suggestion that androstadienone might have a sexual function, and be able to elicit behavioural responses in women towards men in such a manner as to facilitate partnership formation (Hays 2003; Bensafi et al. 2004a; Lundström & Olsson 2005), but so far specific evidence in support of this suggestion has been lacking (e.g. Lundström & Olsson 2005).

2.0.1 Ecological validity

Claims of commercial manufacturers notwithstanding, it is evident that the chemicals named as potential human pheromones do not function as behavioural releasers in the way that pheromones do in other species. Instead of searching for specific reactions to purported human pheromones, it may be that these chemicals are better described as 'modulators' (Jacob and McClintock 2000) which influence psychological states and, thereby, also influence behaviour in a variety of fashions depending on the situation in which they are experienced, or the accompanying cues. If this is true, then the study of 'human pheromones' within the laboratory may be inherently problematic. Indeed, a leading researcher in the field commented of androstadienone: "If it does not function under normal, nonexperimental social conditions, then it is not a pheromone" (Jacob et al. 2002a, p.282). If chemicals such as androstadienone provoke different effects depending on the context in which they are encountered, then their study under nonexperimental social conditions may be a prerequisite for the understanding of their effects. From a theoretical perspective, it is well-established that the co-occurrence of different biological cues can affect their interpretation (Rowe 1999). Yet there are also empirical data that suggest that a consideration of ecological validity may be particularly important in the study of androstadienone.

Researchers have demonstrated sexually dimorphic response to androstadienone at physiological (Bensafi et al. 2004b; Boulkroune et al. 2007; Chopra et al. 2008), electrochemical (Monti-Bloch & Grosser 1991) and affective (Jacob & McClintock 2000) levels. This finding is consistent with the suggestion that androstadienone has a function in sexual signalling, demanding a mate choice context where sex-specific effects are tested. The importance of context has been supported further by a study that showed mood effects of androstadienone to depend on the context that was evoked by film segments (Bensafi et al. 2004a). Finally, research has shown physiological changes and positive mood modulation in women in response to androstadienone when tested by a male but not a female researcher, possibly because a man's presence is necessary to provide the appropriate ecological validity (Jacob et al. 2001a; Lundström & Olsson 2005).

In the study reported below, a "speed-dating" event was chosen as an ecologically valid and theoretically appropriate context to test the effects of potential pheromones involved in human mating behaviour. Speed-dating is a relatively recent form of organised social introduction. Single males and females subscribe to an independently administered event, which is organised to allow each person to interact with each participant of the opposite sex, in a face-to-face meeting, for a pre-defined, limited time period. At the end of each interaction, men and women note covertly whether they would like to meet again. If both parties select the other, their contact details are exchanged by the organisers. Since the aim of participants is to evaluate and attract potential partners, speed-dating provides an ecologically valid and theoretically appropriate context to assess mate choice behaviours (Finkel et al. 2007). The experimental group of women was exposed to androstadienone in a solution smelling of clove, while the two control conditions were made up of a group of women exposed to the clove-scented solution alone, and a group of women in an odourless condition. The speed-dating followed standard conventions, but women were additionally required to rate each man for attractiveness.

2.1 Experiment 1: Speed-dating

2.1.1 Methods

2.1.1 a Participants

The study took place as three separate speed-dating events. Event 1 recruited 22 male and 25 female participants aged 19 – 25 (mean ± SD = 20± 1 yr) from among university undergraduates. To minimise the possibility that participants knew each other, female undergraduates were predominantly recruited from one university and male undergraduates from two separate but nearby universities. The event was held at the students' union and organised by the authors. There was no charge for attendance, and participants received a voucher for a free drink at the end of the evening. The event was advertised as an experiment investigating the effects of odours, including human pheromones, on social judgments. Events 2 and 3 were organised through a private speed-dating agency for participants aged 30 – 45 and 21 – 35, respectively, and took place in a local bar. Female experimental participants were recruited prior to the event by advertisements on the mailing list of the agency, and the usual fee was waived in exchange for their participation. The women were told that they were volunteering for an investigation into the science of attraction, and that they would be exposed to various odours during the evening. At Events 2 and 3, all male and a number of female speed-daters in attendance were not party to the experiment. Data were collected from 17 women aged 35 – 44 (mean ± SD = 39± 3 yrs) who interacted with 19 men in Event 2, and from 12 women aged 21 – 39 (mean ± SD = 32± 4 yrs) who interacted with 25 men in Event 3. All participants were instructed not to consume alcohol before or during the event, or to wear strongly scented products such as perfume or moisturisers.

2.1.1 b Procedure

Participants provided informed consent. They used cotton wool pads to apply either water, clove oil (1% clove oil in propylene glycol), or 4,16-androstadien-3-one (250 μM concentration in the clove oil solution) in sufficient quantities to saturate the area of skin between mouth and nose, from an Eppendorf containing between approximately 0.5 and 1 ml of solution. The presentation method and concentrations were chosen to allow comparison with existing literature (e.g. Jacob & McClintock 2000; Jacob et al. 2001b; Lundström et al. 2003a; Lundström & Olsson 2005). This concentration of androstadienone is thought to be below conscious detection levels for the majority of the population (Jacob & McClintock 2000; Lundström et al. 2003b). Participants were not told which experimental condition they had been allocated to, nor the number of experimental conditions. Before the interactions began, the women answered a few questions (age, self-rated attractiveness, menstrual cycle dates, hormonal contraceptive usage, whether they smoked, whether they had consumed an alcoholic drink in the preceding couple of hours, description of the odour they were given). Three of the women in the androstadienone condition (one from each Event) provided written descriptors of the solution ("urine, babies' nappies", "babies", and "cloves, musty/dirty") that suggested that they may have consciously detected androstadienone; results reported below are qualitatively identical when the ratings given by these women are excluded.

On the basis of existing work showing measurable effects of androstadienone six minutes post-exposure and lasting at least two hours (Jacob & McClintock 2000), there was a gap of at least 15 min between initial odorant application and the beginning of the male-female interactions. Event 1 interactions were completed a maximum of two hours and 15 min after first exposure. In Events 2 and 3, odorants were re-applied during a break approximately two hours after the initial application, and the interactions continued for approximately another hour following re-application. Women were seated at numbered tables, and men moved from table to table at regular intervals as prompted by the organisers. Participants had approximately three minutes to interact with and assess their partner. Female participants were provided with a score card on which they recorded a rating of each man's attractiveness on a scale of 1 – 7, and indicated whether they would like to meet the man again (on the basis that if both participants selected the other, their contact details would be exchanged by the organisers). Photographs were taken of all female participants, and these were rated for attractiveness subsequent to the event by a panel of 15 (Event 1) or five (Events 2 and 3) male raters who had been chosen to be of similar age to those being rated.

2.1.1 c Analysis

For the attractiveness rating scores, scores awarded to one man who left prematurely and so was rated by only five women were discounted; as were all scores given by one woman who did not adhere to the rating system. A further 45 out of a possible 1190 ratings were excluded either because one or both of the participants indicated that they knew the other prior to the event, or because a rating had been omitted from the score card. A maximum of four scores awarded by any one woman, and a maximum of five scores awarded to any one man, were excluded in this way. These 45 omitted or excluded ratings were replaced by the average score given to the man in question, calculated from the remaining valid data. This is a conservative approach, reducing between-conditions variance.

Each man's selection rate was also calculated; this was the proportion of women who selected a man as someone she might want to meet again, divided by the number of women who met that man. The man who left prematurely was excluded. Of the remaining men, two left somewhat early and so failed to meet four women each, and some potential meetings were missed (e.g. if a participant took a break; total omitted meetings n = 15 out of a possible 1173 meetings); the divisor in the calculation was adjusted as necessary.

Analyses were conducted with SPSS version 15.0. Mixed-model ANOVA with man (n = 66) as the unit of analysis and mean attractiveness rating or selection rate for each man calculated from all women in each of the three experimental Conditions (water, clove, androstadienone) as the three within-subject levels, following the analyses of Roberts et al. (2005b) and Wedekind et al. (1995). This analysis may increase the statistical power to detect an effect in contrast to a by-rater analysis, since a by-rater analysis allows only one score to be considered per woman (i.e. 53 scores in the analysis), whereas a by-target analysis enables the consideration of scores to all 66 men as rated by women in the three conditions (i.e. 66 x 3 = 198 scores). Selection rate data were non-normally distributed, but ANOVA is robust to deviations from normal distribution (Subrahmaniam et al. 1975). Greenhouse-Geisser correction was used in the mixed model analyses where data violated assumptions of sphericity. Event (whether the man was rated in Event 1, 2 or 3) was included as a between-subjects factor, although results are qualitatively identical if Event is excluded. Wilcoxon's signed-ranks tests (*T*) were used in place of paired-samples t-tests (*t*), Kruskal-Wallis tests (*H*) used in place of one-way ANOVA, and Kendall's tau correlations (τ) in place of Pearson correlations, if data were non-normally distributed (Kolmogorov-Smirnov, $p < .05$).

2.1.2 Results

2.1.2 a Attractiveness ratings

Experimental Condition (water, clove, androstadienone) had a significant effect on the rated attractiveness of the men ($F_{2,126} = 5.70$, $p = .004$). Men were rated significantly more attractive by women in the androstadienone condition compared with both women in the clove (Wilcoxon's signed-ranks tests, $T = 606$, $p = .015$) and water ($T = 619.5$, $p = .003$) conditions. Ratings from the women in the clove and water conditions did not differ significantly ($T = 961.5$, $p = .468$).

The effect of Event on attractiveness rating bordered on significance ($F_{2,63} = 3.105$, $p = .052$): the trend was for men in Event 1 to be awarded higher ratings, although post-hoc comparisons of the ratings given at each Event revealed no significant differences (Games-Howell procedure: Event 1 and Event 2: $p = .1$; Event 1 and Event 3: $p = .12$; Event 2 and Event 3: $p = .99$).

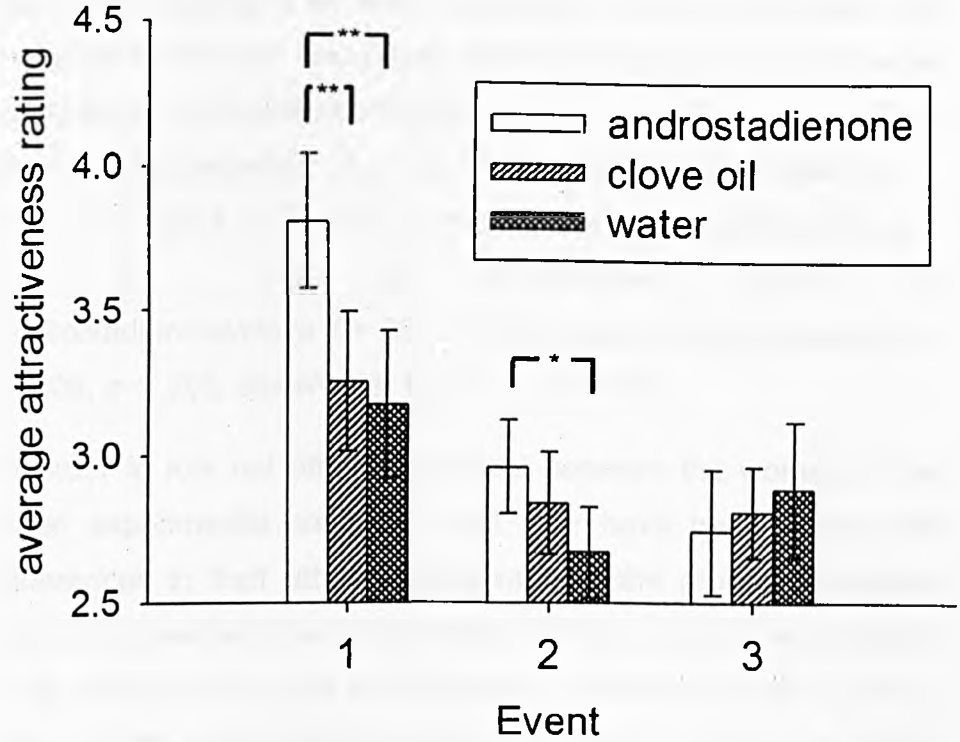

Figure 1. Mean attractiveness rating, by experimental Condition (androstadienone in clove oil, clove oil alone, water) and by Event (speed-dating event 1, 2 or 3). Mean ± SE. ** p < .001 * p < .05.

The main effects of Condition and Event were modified by a significant interaction between Condition and Event ($F_{4,126}$ = 4.89, p = .001; Figure 1), which was analysed further. In Event 1, Condition was significant (repeated-measures ANOVA of Event 1 as above, $F_{2,42}$ = 20.69, p < .001). Men were given higher ratings by women exposed to androstadienone than women exposed to clove (paired-samples t-tests, t_{21} = 5.76, p < .001) or water (t_{21} = 5.41, p < .001), but the ratings from the women in the water and clove conditions did not differ (t_{21} = .73, p = .472). In Event 2, Condition approached significance ($F_{2,36}$ = 3.01, p = .062). The mean rating awarded in the androstadienone condition was higher than that awarded in the clove condition, which itself was higher than that awarded in the water condition, but the only statistically significant difference was between the androstadienone and water conditions (androstadienone/clove (t_{18} = .99, p = .336); androstadienone/water (t_{18} = 2.18, p = .043); clove/water (t_{18} = 1.76, p = .095)). In Event 3, Condition was not significant ($F_{2,48}$ = .384, p = .683; for completeness, figures are: androstadienone/clove T = 89, p = .564; androstadienone/water t_{24} = 1.09, p = .286; clove/water t_{24} = .41, p = .683).

In order to rule out other differences between the women in the three experimental conditions that may have been driving the differences in their attractiveness ratings, the groups of women were compared with respect to their self-rated attractiveness (Little et al. 2001), other-rated attractiveness (Penton-Voak et al. 2003), age, alcohol consumption (Jones et al. 2003; Parker et al. 2008) and (in normally-cycling, non contraceptive users) menstrual cycle phase (Gangestad et al. 2004; Danel & Pawlowski 2006). The women in the three Conditions did not differ significantly in any of these factors (self-rated attractiveness (H_2 = .06, p = .970); other-rated attractiveness (F_2 = .52, p = .597); age (H_2 = .15, p = .929); whether they had had a drink in the preceding couple of hours (H_2 = 2.03, p = .362). The number of normally-cycling women in the 'fertile' phase of the menstrual cycle, 14 – 20 days before predicted onset of next menses, bordered on significance (H_2 = 5.68, p = .056), but while it might be expected that fertile-phase women should give higher attractiveness ratings irrespective of chemical exposure (Gangestad et al. 2004; Danel & Pawlowski 2006; but see Thorne et al. 2002 where no effect of menstrual cycle phase on attractiveness ratings was found), in fact the androstadienone Condition contained the smallest number of normally-cycling, fertile-phase women (one of the 20 women exposed to androstadienone, two of 17 to clove, five of 15 to water).

There was no significant correlation between androstadienone-related attractiveness enhancement (i.e. mean score awarded by women in clove condition subtracted from mean score awarded by women in androstadienone condition) and 'original' attractiveness (i.e. mean score by women in clove condition) (Kendall's $\tau = -.045$, $p = .599$).

2.1.2 b Selection rates

Rated male physical attractiveness co-varied strongly with his chance of being selected for further meetings ($\tau = .741$, $p < .001$; Figure 2).

There was no significant main effect of Condition on selection rates ($F_{1.8, 115.0} = .68$, $p = .495$). Selection rates differed significantly according to Event ($F_{2,63} = 3.16$, $p = .049$), although post-hoc comparisons of the ratings given at each Event revealed no significant differences (Games-Howell procedure: Event 1 and Event 2: $p = .096$; Event 1 and Event 3: $p = .137$; Event 2 and Event 3: $p = .997$). These findings were modified by a significant interaction between Condition and Event ($F_{3.7, 115.0} = 2.95$, $p = .027$; Figure 3). Separate repeated-measures analysis of each Event revealed a significant effect of Condition in Event 1 ($F_{2,42} = 5.14$, $p = .01$); men were selected significantly less frequently by the women exposed to water than the women exposed to clove ($T= 7$, $p = .005$) or to androstadienone ($t_{21} = 2.97$, $p = .007$), while clove and androstadienone conditions did not differ in selection rates ($t_{21} = .22$, $p = .827$). There was no significant effect of

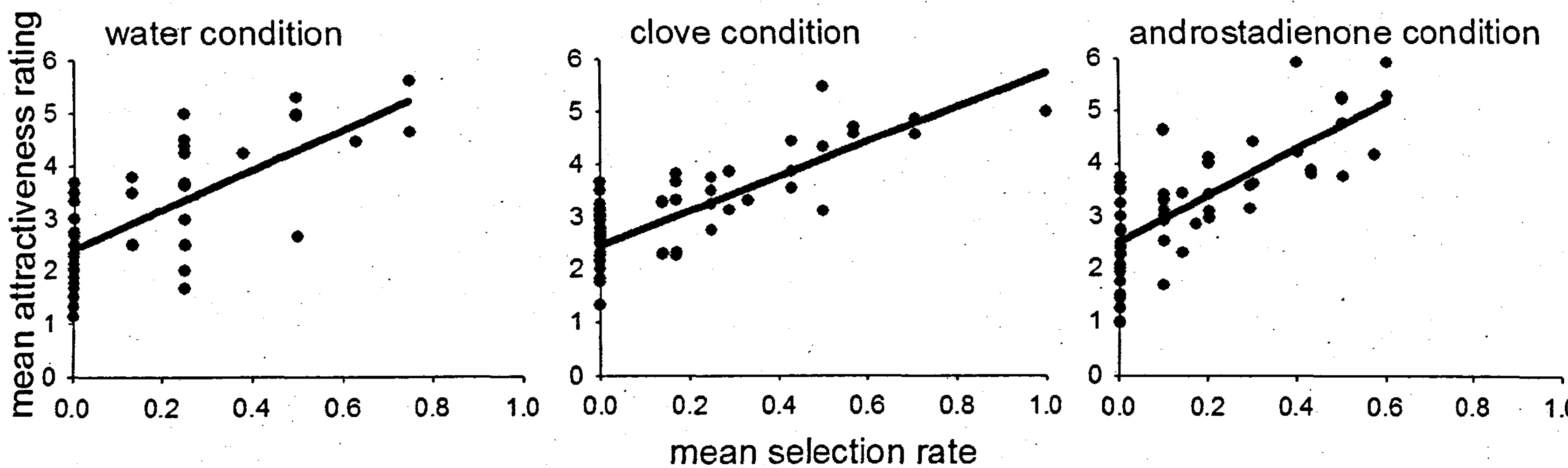

Figure 2. The relationship between the proportion of women who selected a man for possible future meetings and mean attractiveness rating given to that man, by Condition. Each data point represents one man

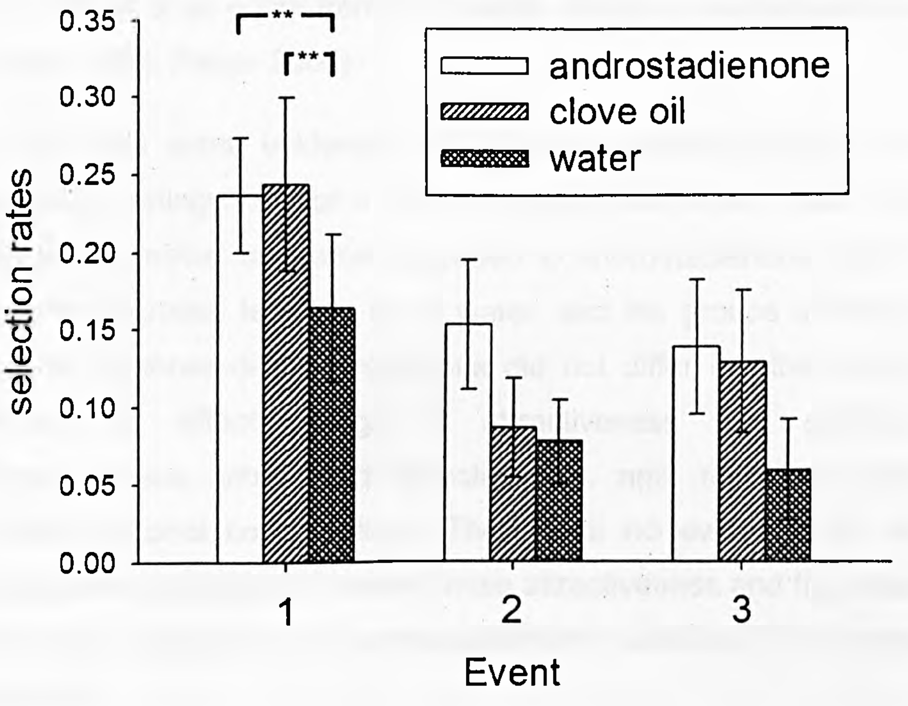

Figure 3. Mean selection rate, by Condition (androstadienone in clove oil, clove oil alone, water) and by Event (speed-dating Event 1, 2 or 3). Mean ± SE. ** p < .01

Condition on selection rates in Event 2 ($F_{2,36}$ = 1.87, p = .169) or Event 3 ($F_{1.6,38.6}$ = 1.62, p = .213).

2.1.3 Discussion

The investigation examined the effects of androstadienone exposure on women's attractiveness ratings, and selection, of men, at three separate speed-dating events, in a between-subjects design. Speed-dating was chosen as a suitable arena to investigate a chemical with a posited sexual signalling function. Women applied either androstadienone mixed in a solution that smelt of clove oil, the clove oil solution alone, or water, to the upper lip, enabling passive inhalation of the solution throughout the evening. The three experimental conditions allowed us to parse the effects of an odour from the specific effects of androstadienone (Hays 2003; Pause 2004).

There was some evidence that androstadienone exposure can enhance ratings of men's attractiveness. Men were rated more attractive overall by women exposed to androstadienone than by women exposed to clove oil or water, and the groups of women given the three different solutions did not differ in other factors known to affect ratings of attractiveness (i.e. self-rated attractiveness, other-rated attractiveness, age, menstrual cycle phase, alcohol consumption). There was no evidence for any significant relationship between male attractiveness and the extent to which women in the androstadienone condition gave higher ratings.

However, the main effect of androstadienone exposure was modified by a significant interaction with Event (i.e. at which of the three separate speed-dating events the data were collected). The most compelling evidence for the effects of androstadienone came from the first event, with some confirmatory evidence from the second event, and no demonstrable effect of androstadienone at the third. The reasons for this interaction are unclear. Sampling effects could be at issue: the first event contained the most, and the third the fewest, female participants. The effects of androstadienone appear to be amenable to context, and it is likewise possible that differences between the speed-dating events were influential. The first event, organised by the authors, was most carefully controlled, and participants may have been more motivated; certainly, at the first event there was a trend for overall ratings of attractiveness and selection rates to be higher.

Alternatively, a greater proportion of the women in the first event may have been sensitive to the effects of androstadienone. The composition of the people in the studies may also play an important role. The first event contained the youngest sample, where physical traits may be relatively more important than, for example, financial status, compared to older samples, and this could increase the chances that odour cues might be important. Results may also be prone to variation amongst the attractiveness of the people participating in each event.

Although exposure to sub-threshold concentrations of non-odorous sweat (but not synthetic, odorous sweat, Demattè et al. 2007) has been shown to enhance women's ratings of men's attractiveness (Thorne et al. 2002), a previous experiment, in which the effect of androstadienone on ratings of facial photographs was measured, detected no significant response (Lundström & Olsson 2005). If androstadienone can modulate social perceptions, it is possible that ecologically valid accompanying cues, such as a man's presence, are necessary for this effect to emerge (Jacob et al. 2001a; Lundström & Olsson 2005; Saxton et al. 2007).

The finding that androstadienone may enhance women's ratings of men's attractiveness is consistent with the idea that it may constitute a sexual signal that conveys information about male mate quality. The finding that women show concordant strength of preference for male facial masculinity and the odour of androstadienone has been interpreted as indirect evidence that androstadienone may index male mate quality (Cornwell et al. 2004). In male boars, production levels of a related chemical, a boar pheromone called androstenone, correlate positively with social rank and aggression levels (Giersing et al. 2000).

Although it is assumed that the results are due to an effect of androstadienone on the women, it is not inconceivable that the men were also behaving differently in its presence, despite their greater distance to the source of the odour. Certainly, male responses to androstadienone have been noted previously; these are often different from female responses (Jacob & McClintock 2000; Bensafi et al. 2004b; Boulkroune et al. 2007; Chopra et al. 2008).

An understanding of the effects of androstadienone is still far from complete, and this required a number of assumptions to be made in the experimental design. Firstly, although an experimental context that was more ecologically valid than a standard laboratory test was used, the method of application of androstadienone, and possibly also the dosage, remains somewhat artificial (Saxton et al. 2007), and, as indicated in the results, the androstadienone concentrations used may be consciously detectable by a minority of the population (Jacob & McClintock 2000; Lundström et al. 2003b). It is not clear what would constitute an appropriate quantity of androstadienone, to reflect that which may be experienced during normal non-intimate social contact (Wysocki & Preti 2004; Lundström & Olsson 2005). Analysis of androstadienone concentrations within the apocrine sweat of a small sample of individuals has revealed values as high as 1900 µM (Gower et al. 1994), although analysis of the quantities of androstadienone deposited upon axillary hair (excluding the skin) over a 24-hour period resulted in much lower quantities, up to just over 4000 pmoles of androstadienone (Nixon et al. 1988). The quantities that might reach a social partner are unknown. It was decided to use an experimental solution containing androstadienone at a concentration of 250 µM, to be applied to the region between the upper lip and the nostrils, in order to replicate the concentration and application method commonly used in such experiments (Jacob & McClintock 2000; Jacob et al. 2001a; Lundström & Olsson 2005), thereby allowing for comparison with existing literature. Androstadienone sensitivity can vary widely; since the concentration used in the study is close to the average odour threshold of the population (Lundström et al. 2003b), the convention was followed of masking its odour in strongly-scented clove oil (1% clove oil in propylene glycol). It is debatable whether something may be classified as a 'pheromone' if it is consciously detected, and so this procedure aimed to prevent conscious or learnt associations with the odour from influencing its effects. The participants were exposed to the solutions by application to the upper lip, from where they were able to inhale the solutions throughout the evening. This does not exclude the possibility that the solutions were transmitted dermally rather than by odour. However, another study has noted mood effects of androstadienone which were not distinguishable according to whether exposure to the chemical was enabled by upper-lip application or passive inhalation (Jacob et al. 2002a), and indeed the method of transmission is of little consequence to the current investigation.

An alternative design which may increase ecological validity further would be to apply the odour stimuli to the men, rather than to the women; and to use sub-threshold concentrations without the clove oil admixture. The use of clove oil as a masking odour allows for comparison with previous literature, and its use as an odour control is critical for distinguishing specificity of effect. Its usage has been criticised on the basis that the experimental solution then constitutes a complex mixture (Lundström et al. 2003a), and also that it has a number of biological effects including anaesthesia (Chaieb et al. 2007), but since it was used in both the experimental and control conditions, any possible biological effects of clove oil should not affect the experimental findings. Related to these problems, it is as yet unclear whether androstadienone can be expected to have any effect in naturalistic settings other than in very intimate dyadic encounters, or if it does, how a man might solve the problem of influencing a woman's response to him without influencing her response to all other men in the vicinity.

Our results represent a novel demonstration of the modulation of mate choice behaviour. Yet they are also consistent with a range of findings that may constitute the proximate mechanisms of the modulation: enhancement of female positive mood (Jacob & McClintock 2000; Jacob et al. 2001a; Bensafi et al. 2004a; Bensafi et al. 2004b; Lundström & Olsson 2005; Villemure & Bushnell 2007); or increased ease-of-processing of facial characteristics in the presence of related cues such as bodily odours (Rowe 1999; Kovacs et al. 2004). While it is possible that attractiveness modulation is a mere corollary of these previously noted effects, it could as easily be argued that modulation of reproductive behaviour is the primary function. The studies are highly suggestive that the presence of androstadienone can increase a woman's attraction to a man in a mate choice context, though the strength of this effect and the exact context in which it applies remain questions for future research. The next chapter goes on to examine whether attractiveness in the visual and olfactory channels are equally affected. It also returns to the question of whether, if androstadienone enhances attractiveness, it interacts with a more objective measurement of male quality.

Chapter 3: Interactions with androstadienone

In Chapter 2, an experiment was reported that contrasted groups of women exposed to either water, clove oil, or androstadienone in clove oil at a speed-dating event where the women rated each male for attractiveness. The group of women who had been exposed to androstadienone gave significantly higher ratings than the other two groups, suggesting that androstadienone may be associated with perception modulation related to mate choice. However, a similar experiment by Lundström et al (2005) in which women were exposed to either clove oil, or androstadienone in clove oil, in a within-subjects experiment, found no effect of condition on the ratings of attractiveness of a set of male facial photographs. The reason for the discrepancy may be that the effects of androstadienone on person perception are holistic rather than specifically visual, and are better detectable in ecologically-valid contexts (Jacob et al. 2002a; Saxton et al. 2007). Consistent with this, effects of androstadienone have been detected when participants have been tested by a male but not a female researcher, leading researchers to suggest that the presence of a male researcher may be necessary to add appropriate ecological validity (Jacob et al. 2001a; Lundström & Olsson 2005). Androstadienone, as a chemical of human origin, would normally be experienced only in social contexts, and may be most relevant in relation to opposite-sex interactions. Alternatively, there may have been differences in susceptibility to androstadienone in the samples; androstadienone detection varies widely in a population (Lundström et al. 2003b), and these differences may result from both individual differences in experience (Chopra et al. 2008) and individual differences in genetic make-up (Keller et al. 2007). In light of open questions such as these, the scope of perception modulation by androstadienone merits further investigation. In this chapter, two experiments that examine perceptions in the olfactory and visual channels are reported.

3.0.1 Visual and olfactory assessment

The visual assessment of facial attractiveness from photographs is a widely used tool in human mate choice research (see e.g. Penton-Voak et al. 1999b; Rhodes 2006). Facial perception is supported by dedicated neural areas (Kanwisher & Yovel 2006), and even neonates discriminate face-like objects (Goren et al. 1975; Johnson et al. 1991). Human judgments of facial attractiveness tend to be broadly similar even between different ages, sexes and cultures (Langlois et al. 2000), and face-based judgments of attractiveness, likeability, trustworthiness, competence, and aggressiveness made in as little as 100 ms correlate highly with judgments made in the absence of time constraints (Willis & Todorov 2006). In light of all this, differences between raters in terms of their ratings of facial attractiveness has been commonly used as evidence in support of systematic differences between the raters in terms of their mate preferences (Penton-Voak et al. 2004; e.g. Jones et al. 2005b; Roberts et al. 2005b; Saxton et al. 2006). Ratings of facial attractiveness appear particularly amenable to modulation by the immediate context, including simultaneously-perceived odours (Thorne et al. 2002; Demattè et al. 2007), and exposure to other faces of varying levels of attractiveness (Little & Mannion 2006). However, particularly for female raters, the olfactory quality of a mate or potential mate may vie for importance with the visual qualities (Herz & Cahill 1997; Herz & Inzlicht 2002; Havlicek et al. 2008). Furthermore, since androstadienone is perceived in the olfactory channel, investigation of the modulation of odour perception in its presence seems apposite, and yet has so far gone unexplored. The first experiment thus set out to explore whether androstadienone could modulate perceptions of both body odours and accompanying facial photographs. If ecological validity is necessary to evoke effects of androstadienone, the presence of odour cues might create the sufficient ecological validity that may be missing from facial photographs alone in the study by Lundström & Olsson (2005).

However, even if androstadienone modulates perceptions of attractiveness, it is possible that this modulation does not affect all stimuli equally; it may interact with some measure of the reproductive quality of the stimulus. In Chapter 2, we saw no evidence for a correlation between a man's attractiveness, that is, greater or lesser mate quality, and the extent to which he benefited from the modulation. A more objective measure of reproductive quality may be MHC heterozygosity. As outlined in Chapter 1, the MHC (or HLA in humans) is the set of genes which code immune response. MHC variation has been linked to reproductive quality and mate desirability in humans as in other animals (for reviews see e.g. Penn & Potts 1998; Penn 2002; Roberts & Little 2008). Greater levels of heterozygosity have been associated with more robust immune system function (Penn & Potts 1998; Penn 2002) and also with rated attractiveness (Roberts et al. 2005c). The second experiment thus examined whether differences in attractiveness and MHC heterozygosity are associated with differences in the extent to which males benefit from any attractiveness modulation by androstadienone.

3.1 Experiment 3a: Body odour and androstadienone

3.1.1 Methods

7 men (aged 20 – 30) were recruited from amongst university undergraduates and provided informed consent. Each man was supplied with two sterile cotton pads (Asda) and tape for dressings (Micropore) together with full instructions. The day before sample collection, men were instructed to shower with fragrance-free soap (Asda), not to apply deodorant, and to tape a pad under each axilla. Pads were worn for up to 24 hours, during which time men were instructed to avoid intimate contact with others. Upon removal, men placed the pads in the ziplock plastic food bags supplied. Cotton pads were worn for a length of time between 18.5 and 24 h, and there was a time period of no more than 2 h 50 mins between pad removal and transfer to cold storage. At the time of sample collection, men supplied details of their age, were photographed (Canon Powershot) under artificial light against a neutral background, and were questioned to confirm that they had followed sample collection instructions. Pads were stored at approximately -75° C for 17 h overnight, and removed from the freezer 1.5 h before odour assessment began. This method of odour collection and storage has previously been used successfully in body odour perception research (e.g. Roberts et al. 2005a).

To ensure that samples in each experimental condition were not influenced by possible differences in body odour production between right and left axillae (Hopwood et al. 2005), each pad was cut into two pieces, creating four pieces from each male donor. Two pieces from each man (one from the left and one from the right axilla) were allocated to the androstadienone condition and two to the control condition. Samples from different men were not combined. 0.3 ml of propylene glycol alone was applied to the pads in the control condition. 0.3 ml of a solution of 0.25 mM androstadienone in propylene glycol (a conventional experimental concentration; Jacob & McClintock 2000; Jacob et al. 2001b; Lundström & Olsson 2005) was applied to the pads in the androstadienone condition. Solutions were added to the samples 15 mins before rating began. Samples from one male donor were used for the pre-experimental exposure (see below) and the remaining six men were used for the experiment.

20 women (aged 20 – 45) were recruited from among university students. Women were told that they would need to attend two sessions where they would be presented with body odour samples and photographs from a set of men, and asked to rate these stimuli for attractiveness. Women provided informed consent and were debriefed after the experiment. They were not told the experiment hypotheses. All rating took place on one day. Experimental condition order was counterbalanced across women and double-blind. Exposure to the pads in the androstadienone condition was separated from exposure to the pads in the control condition by at least 2 ½ hours to attempt to avoid transfer effects between conditions (Jacob & McClintock 2000). To allow for the development of the response (Jacob & McClintock 2000), each woman was asked to sniff a sample from the same male donor from the appropriate condition (control or androstadienone) at least five minutes in advance of each testing session, under the pretext

of ensuring that she could detect the types of odours to be used. All women, apart from one prior to her androstadienone session due to her non-availability, were exposed to the odour from the appropriate condition in this way between five and 30 minutes before each testing session. The odour used in the initial exposure was not re-used in the testing sessions.

Testing took place in a well aerated room. Pads were placed in numbered 500 ml pyrex conical flasks with foil caps which could be removed to sample the odour and then replaced. Photographs were presented in separate numbered envelopes. Ratings were carried out on a seven point scale anchored by verbal descriptions. First, women sniffed the jars and rated (1) the attractiveness and (2) the intensity of the body odour in each jar. Second, the women opened the envelopes and rated the attractiveness of each man for (3) a short-term and (4) a long-term relationship. Lastly, the women were asked to pair the numbered jars and photographs and assess both facial photograph and body odour together in order to rate each man for (5) a short-term and (6) a long-term relationship. For each rating, women also had the opportunity of indicating on their rating sheet that they could not detect anything which smelt like body odour, or that they recognised the male in the photograph: in either of these instances, they were not expected to rate that stimulus. For the second session, women were told that they might have seen some of the males in the photographs in the first session, but that they should approach the task afresh. Up to three women attended any individual rating session at one time.

The mean rating given in each condition to each question by each woman was calculated. If a woman indicated that she could not detect anything that smelt like body odour in either of the sessions, the rating she gave to that sample was excluded from both sessions. There were four occurrences of a sample being deemed odourless. If a woman indicated that she recognised the man at any point during the sessions, all photograph ratings (i.e. questions 3 – 6) were excluded from both sessions. One man was recognised by two women. Exclusions thus affected 48 out of a possible 1440 ratings. Excluded scores were replaced by the average score of all of the other women. This is a conservative method, reducing between-conditions variance. Data analysis was carried out in SPSS 15.0. Rater as unit of analysis allowed consideration of all six dependent variables (i.e. from all six scores) in one analysis. Greenhouse-Geisser correction was used when data were non-spherical (Mauchly's test of sphericity, $p < .05$). Differences between the sets of scores used in the paired-samples t-test were normally distributed (all Shapiro-Wilk $p > .1$).

3.1.2 Results

Repeated-measures analysis of mean ratings for each question (within subjects factors: question 1 – 6, experimental condition) showed that women gave significantly higher ratings in the session when they were exposed to androstadienone compared with the control condition ($F_{1,19} = 5.96$, $p = .025$) with a medium effect size ($r = 0.47$)(Figure 4). There was also a significant effect of question ($F_{2.5,48.2}$ 20.027, $p < .001$), but no interaction between condition and question ($F_{2.0,37.9} = .563$, $p = .728$). The inclusion of condition order (whether the androstadienone condition was first or second) in the model did not qualitatively change the results, and there was no main effect of or interaction with condition order (all $p > .24$).

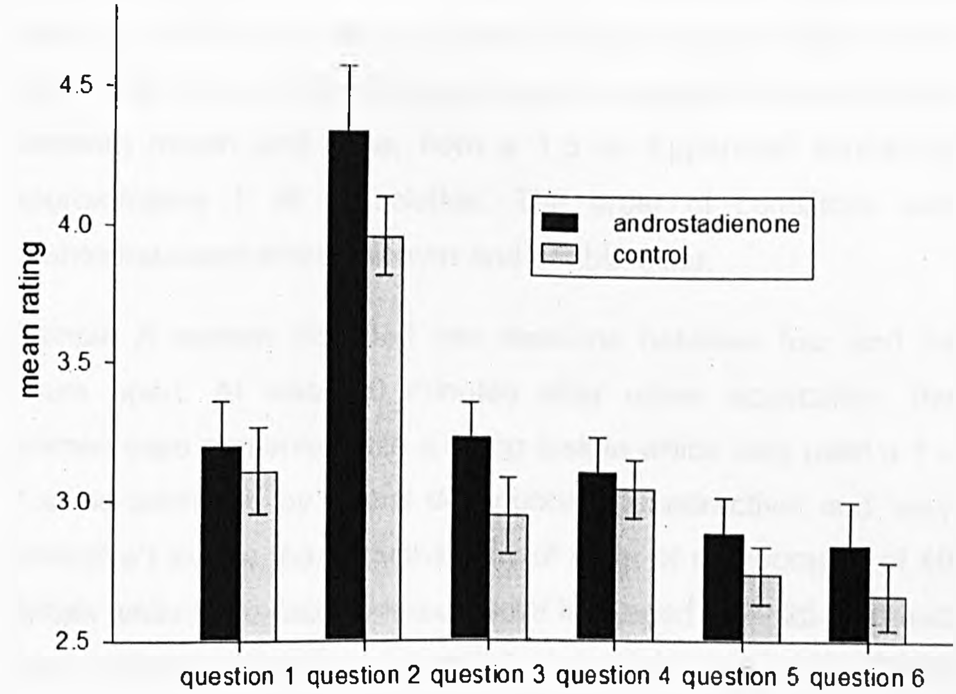

Figure 4. Mean ratings by condition. Question (1) body odour attractiveness and (2) intensity; photograph attractiveness for a (3) short-term relationship and (4) long-term relationship; overall attractiveness for a (5) short-term and (6) long-term relationship.

3.2 Experiment 3b: Interactions with quality

3.2.1 Methods

Data were collected in two rounds (Samples A and B). Sample A consisted of 20 women and Sample B 19 women aged 18 – 35, all of whom reported themselves to be predominantly or exclusively heterosexual, who were recruited from among university students and social contacts and consented to participate in a two-session study on the effects of odour on perceptions. At each session, women used cotton wool pads to apply conventional experimental concentrations (Jacob et al. 2001b; Lundström & Olsson 2005;

Saxton et al. 2008) of either clove oil (1% clove oil in propylene glycol) or 4,16-androstadien-3-one (250 μM concentration in the clove oil solution) in sufficient quantities to saturate the area of skin between mouth and nose, from a 1.5 ml Eppendorf containing approximately 1 ml of solution. The order of conditions was counterbalanced across women and double-blind.

Sample A women attended two sessions between four and 24 hours apart. At least 10 minutes after odour application, the women were presented with a rating task in which they used a 1 – 7 scale anchored by verbal descriptions ('unattractive' and 'very attractive') to rate the attractiveness of a set of photographs of 40 British white male faces without facial hair aged 18 – 25 who had been genotyped at three key MHC loci, HLA-*A*, -*B*, and –*DRB1* (full details at Roberts et al. 2005c). 20 of the men were heterozygous at all three loci and 20 of the men were homozygous at one (n = 16) or two loci. Sample B women carried out the rating task between 40 and 60 minutes after odour application, following an unrelated task. Sample B session scheduling was constrained by female and testing room availability, but days between sessions were approximately counterbalanced; sessions took place one or two (n = 6 androstadienone, 7 clove condition first) or between five and 14 (n = 3 androstadienone, 3 clove condition first) days apart. Women were debriefed by email after the experimental sessions with all of the participants had been completed.

After the second session, Sample A women completed a questionnaire that asked whether the odour smelt odd or musky; all replied in the negative. After the second session, Sample B women were asked whether the odour smelt differently from the previous session. If a woman replied affirmatively (n = 10), she was given an odour sensitivity test in which she was presented with five sets of three 1.5 ml Eppendorfs and asked to note down the odd one out from each set of three. Of each set, two contained 1 ml of the clove oil solution and one contained 1 ml of the androstadienone in clove oil solution. Women were left alone to complete this task. Six women correctly selected the Eppendorf containing the androstadienone solution as the odd one out three or more times out of five; results are qualitatively identical with the exclusion of these six women.

Menstrual cycle stage and hormonal contraceptive usage (Sample A, n = 10 non-users; Sample B, n = 18 non-users) were noted because this may affect ratings of facial attractiveness (review in Jones et al. 2008a). Onset of next menses of Sample B women was confirmed by email subsequent to the sessions. Amongst Sample B women, two sessions (*n* = two women; one androstadienone condition, one clove condition) were held during the 'fertile window' (Wilcox et al. 1995) 15 to 19 days prior to next onset of menses (i.e. Day 10 – 14 of a 28-day cycle), where counting backwards from next onset of menses to determine fertility is more accurate than counting forwards from last menses (see Johnston et al. 2001). Sample A women carried out the rating tasks at most 24 hours apart, thereby minimising any effect of menstrual cycle phase.

Analysis was carried out in SPSS 15.0. Rater as unit of analysis allowed usage of the repeated-measures format in its traditional sense, i.e., to compare scores across time. Non-parametric correlations were used when data were non-normally distributed (Shapiro-Wilk, $p < .05$). There was some indication of non-normal distribution of the differences between the average ratings given by each woman in the two conditions (Shapiro-Wilk p = .015 but Kolmogorov-Smirnov $p > .2$); results are qualitatively identical with the non-parametric Wilcoxon signed-ranks test. Sample was included as a factor since there were differences between the two groups, including length of time between sessions, hormonal contraceptive usage, and completion of an unrelated task prior to the testing by the Sample B women.

3.2.2 Results

A mixed-model ANOVA comparing the mean attractiveness rating given by each woman in the two conditions (between-subject factors: Sample A or B; within-subject factors: Condition) revealed no main effect of Condition ($F_{1,37}$ = 1.591, p = .215). Women in Sample B gave significantly higher ratings than women in Sample A ($F_{1,37}$ = 72.407, $p < .001$). There was no interaction between Condition and Sample ($F_{1,37}$ = 1.348, p = .253). Results are qualitatively identical if hormonal contraceptive usage is included as a between-subjects factor.

Following the analyses of Roberts et al. (2005b) and Wedekind et al. (1995), examination was also made to determine whether men were given higher ratings in the two conditions in a paired-sample t-test where man was unit of analysis. There was no effect of Condition on ratings ($F_{1,39}$ = .071, p = .792); results were qualitatively identical with the inclusion of heterozygosity as a within-subjects factor ($F_{1,38}$ = .072, p = .790). Since there were significant differences between the ratings by Sample A and Sample B women that could not be statistically controlled when man was unit of analysis, the Samples were also analysed separately. There was no effect of Condition on Sample A ratings (t_{39} = .247, p = .806), but men were given significantly higher ratings by Sample B women in the androstadienone compared with the control condition ($t_{39} = 2.299$, $p = .029$).

Since there was some evidence that women in Sample B were giving higher ratings in the androstadienone condition, the next analysis considered whether there was any interaction between the target male and the rating increase for Sample B alone. There was no relationship between the extent to which a male was rated more attractive in the androstadienone compared to the clove condition and the rating of his attractiveness by a prior group of 36 British white women aged 18 – 34 (full details in Roberts et al. 2005c) (Kendall's $\tau = -.046$, $p = .675$), and no effect of heterozygosity on attractiveness increase (mixed model ANOVA; within-subjects factors: androstadienone or clove condition; between-subjects factors: heterozygous or homozygous at the MHC; $F_{1,38} = 1.977$, $p = .168$; Condition remains significant).

Finally, the association between androstadienone and increased discrimination between heterozygosity and homozygosity was examined. For each woman, the mean score that she gave to the heterozygous men was divided by the mean score that she gave to the homozygous men in each condition. Ratios were significantly more equal in the androstadienone condition (mixed model ANOVA; within-subjects factor: condition; between-subjects factor: Sample; $F_{1,37} = 5.834$, $p = .021$), with no effect of or interaction with Sample (both $p > .2$).

3.2.3 Discussion

In Experiment 3a women rated body odour samples, facial photographs, and the overall attractiveness of six males in a within-subjects design in two conditions. In one condition, androstadienone in a carrier solution of propylene glycol had been added to the body odour samples; in the other condition, the carrier solution alone was added. Women gave significantly higher ratings in the androstadienone condition, with no interaction between the condition and the type of rating (e.g. body odour attractiveness). That is, attractiveness enhancement was not restricted to one aspect of male attractiveness, but influenced ratings of odour, visual attractiveness, and overall attractiveness.

In Experiment 3b, two samples of women rated the attractiveness of facial photographs of 40 males in a within-subjects design in two conditions where androstadienone in clove oil, or clove oil alone, had been applied to the area of skin below the nostrils to allow for passive inhalation of the odour. The second sample of women gave significantly higher ratings than the first sample irrespective of condition, and there was some evidence that the second sample gave significantly higher ratings in the androstadienone relative to the control condition. However, this was only apparent when man was unit of analysis. This discrepancy may be because two women in particular were driving the effect, one of whom gave unusually low ratings in the control condition and one of whom gave unusually high ratings in the androstadienone condition. Further research is required to determine whether this is a random fluctuation, or a consequence of individual differences in androstadienone detection and response (c.f. Lundström et al. 2003b; Keller et al. 2007).

Together, these results are somewhat consistent with previous findings that androstadienone exposure is associated with the enhancement of attractiveness ratings, especially in more ecologically valid contexts (Chapter 2; Saxton et al. 2008). They are somewhat inconsistent with previous findings that androstadienone does not enhance ratings of photographic attractiveness (Lundström & Olsson 2005). The reason for the discrepancy may be that photographic stimuli are overly sparse replicas of real-life men. Experiment 3a presented better evidence than Experiment 3b for modulation of attractiveness judgments, and only in Experiment 3a was odour was presented alongside visual cues, thereby better reflecting men in real life.

Experiment 3b also sought to determine whether, amongst the second sample of women in which there was some evidence for an influence of androstadienone on attractiveness rating, there was any interaction between rating differences between the two conditions and male mate quality, measured by MHC heterozygosity or by rated attractiveness. There was no relationship between a male's attractiveness and his attractiveness increase, nor any effect of heterozygosity on attractiveness increase. However, when Sample A and Sample B women were considered all together, there were significant differences between the two conditions in terms of women's discrimination, as measured by the set of ratios created by dividing the mean rating that she gave to the heterozygous men by the mean rating that she gave to the homozygous men. Ratings of heterozygous and homozygous men were significantly more equal in the androstadienone condition. This implies that the less reproductively desirable homozygous men benefit more than heterozygous men from the presence of the androstadienone; if androstadienone is a cue of quality, there may be a ceiling effect or an asymptote where more desirable men benefit less. Yet this is not consistent with the finding that only women in the second sample, who gave higher ratings irrespective of condition, may have been influenced by androstadienone; nor with the finding of no correlation between a man's attractiveness and the extent to which he benefited from androstadienone. Accordingly, the results would benefit from replication.

The results also have methodological implications for research on androstadienone. The solution of androstadienone that was added to the body odour samples (Experiment 3a) or applied to the area of skin below the nostrils (Experiment 3b) followed standard experimental concentrations (Jacob & McClintock 2000; Jacob et al. 2001b; Lundström & Olsson 2005); previous research indicates that this concentration may be consciously detectable by approximately 14 % of a sample of male and female students when presented in a 15 ml solution of propylene glycol in 250 ml polypropylene squeeze bottles (Lundström et al. 2003b), and another study found that three of nine women detected a musky smell when androstadienone was presented at this concentration but without clove oil (Jacob & McClintock 2000). Nevertheless, in Experiment 3a, the concentration used appears to have affected odour samples to the extent that they were rated as more intense in the androstadienone condition. In Experiment 3b, despite the use of the clove oil masking odour, over one half of the women of the second sample reported differences between the subjective qualities of the first and second odour, and nearly one third of the women demonstrated the ability to distinguish consciously between the odours in an odd-one-out test. Androstadienone when smelt at above-threshold levels is often evaluated negatively (Boulkroune et al. 2007), and it is possible that a lesser quantity would have had a more positive effect on ratings (Li et al. 2007). Together, these results suggest that caution should be used in employing the standard concentrations which have been used previously in androstadienone research if conscious detection of androstadienone is to be avoided. To avoid this, future research might consider using lower concentrations, a higher concentration of clove oil (Jacob et al. 2002a), or passive inhalation rather than dermal application (Jacob et al. 2002a). The study design also assumes that any naturally-occurring quantities of androstadienone in the body odour samples used were sub-optimal for the modulation of judgments, and that the addition of androstadienone could enhance ratings, although this has not been specifically tested.

The question of the proximate pathway of any judgment modulation by androstadienone remains open. Androstadienone has been previously associated with general female mood enhancement (Grosser et al. 2000; Jacob & McClintock 2000; Jacob et al. 2001a; Lundström et al. 2003a; Bensafi et al. 2004a; Lundström & Olsson 2005; Villemure & Bushnell 2007), an effect that could lead to enhanced social receptivity, including ratings of attractiveness. Alternatively, the simultaneous presentation of multiple cues can lead to enhanced ease of detection and memorability (Møller & Pomiankowski 1993; Rowe 1999; Candolin 2003), an effect that has been demonstrated specifically in relation to facial stimuli presented simultaneously with bodily odours (Kovacs et al. 2004). Increased ease of processing by the simultaneous perception of facial stimuli and an odour, androstadienone, that has corporal associations, could lead to an enhancement of consciously perceived attractiveness (c.f. Winkielman et al. 2006, who seek to explain the appeal of facial averageness with reference to ease of processing).

Experiment 3a sought to distinguish the impact of androstadienone on various modalities by asking subjects to provide separate ratings in relation to each. However, it is not clear that changes in attractiveness perception can be sensibly localised to any particular modality; cross-modal interaction between the senses is well established (e.g. Zuckerman et al. 1991; Castiello et al. 2006). The most parsimonious description may be that female affect towards men became more positive, or even merely that female affect in general became more positive, and that more positive physical evaluations constitute one instantiation of this change. Irrespective of the mental substrate to these changes, the results present evidence for androstadienone-related alteration of mate choice behaviours.

Short-term variation: general discussion

The three experiments in Chapters 2 and 3 presented novel evidence for a link between androstadienone exposure and the enhancement of female ratings of male attractiveness. This expands our understanding of the function of androstadienone, and recommends a number of avenues of further investigation.

Most urgently, the results imply that investigation of a link between male quality and androstadienone production may be productive. The few studies that have investigated androstadienone production levels have demonstrated a large amount of individual variability (Nixon et al. 1988; Fukushima et al. 1991; Gower et al. 1994). Conversion between testosterone and androstadienone has been demonstrated *in vitro* (Stylianou et al. 1961), presenting a possible biological pathway between androstadienone production and mate quality, and supporting indirect evidence for such a link (Cornwell et al. 2004). A study of a relatively small sample of men (n = 11) found a positive but low-effect correlation between androstadienone and testosterone plasma levels; significance values are not given and the authors report the relationship as the absence of any correlation (Fukushima et al. 1991). Further work is merited.

The introductory comments to Section 1 noted two distinct approaches to psychological investigation of how humans are influenced by biologically relevant odours. One approach is to consider the effects of a chemical, irrespective of the information conveyed. This approach has characterised research on chemicals such as androstadienone. A second approach is to privilege the role of the information above the identity of the chemicals involved. This approach is more common amongst investigation of how

MHC type may influence preference, for instance. Investigation of the information, if any, conveyed by androstadienone, would have the benefit of combining methodology from both of these approaches.

This discussion would be incomplete without a comment on the enticing controversy surrounding 'human pheromones'. It is undisputed that non-human animals make use of pheromonal communication (e.g. Wyatt 2003). Yet the notion of 'human pheromones' is controversial (e.g. Ben-Ari 1998; Jacob et al. 2002a; 2005). This is apparent even in terms of the terminology used; in humans, published reports of physiological or psychological responses to substances that are chemically related to animal pheromones refer to these non-committedly as 'steroids' (Bensafi et al. 2004a; Bensafi et al. 2004b; Boulkroune et al. 2007; Villemure & Bushnell 2007), diplomatically as 'chemosignals' (Jacob & McClintock 2000; Jacob et al. 2001a; Jacob et al. 2001b; Lundström et al. 2006; Saxton et al. 2008), or cautiously as 'putative pheromones' (Monti-Bloch & Grosser 1991; Lundström et al. 2003a; Gulyas et al. 2004; Savic et al. 2005). The roots of this debate are in part empirical. A pheromone must function in a normal, non-experimental context (Jacob et al. 2002a). Although the experiments of Chapters 2 and 3 show that this may be possible in relation to androstadienone response, they do not show that the quantities actually emitted have that effect. Yet there seems also to be another component to the objection to the notion of 'human pheromones', which may relate to a prissiness surrounding the conception that humans are subject to the same biological forces that affect other animal species (see e.g. Dennett 1995). Researchers of the behavioural effects of androstadienone seem most comfortable with the term 'chemosignal', and yet it should not be overlooked that this is in some senses a stronger claim than usage of the term 'pheromone'. Pheromones are, "chemicals which are secreted to the outside by an individual and received by a second individual of the same species, in which they release a specific reaction, for example, a definite behaviour or a developmental process" (Karlson & Lüscher 1959, p.55). A signal is, "Any act or structure that (i) affects the behaviour of other organisms; (ii) evolved because of those effects; and (iii) which is effective because the effect (the response) has evolved to be affected by the act or structure" (Scott-Phillips 2008, p.388). Naming a chemical as a chemosignal thus makes a specific claim not only to its function, but also to its phylogenetic history and the nature of the relationship (i.e. collaborative rather than coercive) between the sender and receiver.

The experiments in Section 1, then, showed how humans may adjust their mating preferences on an immediate basis in relation to cues available in the environment. The experiments in Section 2 go on to examine how human mating behaviours may be adjusted over the medium term.

SECTION 2: MEDIUM-TERM VARIATION: EFFECTS OF THE MENSTRUAL CYCLE

Next, in Chapters 4 and 5, we turn to the menstrual cycle. This topic has recently met with much scientific enquiry, with reports that a woman's appearance, and her mate preferences, change systematically across the cycle. The menstrual cycle was selected as the arena in which variation would be investigated across the medium term because the time span is clearly delineated, and because research on this topic is burgeoning. Further, this research topic is well defined both at the proximate level by virtue of the hormonal substrate, and at the ultimate level with reference to the reproductive consequences of the phases of the cycle. Existing literature often embraces both levels, and accordingly these are introduced next.

From the perspective of behaviour as biological adaptation, the most interesting consequence of the menstrual cycle is that it marks the systematic rise and fall in the likelihood of conception following intercourse. Data from a relatively large study of 221 healthy women who were trying to become pregnant showed that, of the 192 pregnancies initiated, conception only took place when intercourse had occurred during a six-day period that ended on the day of ovulation, where ovulation was estimated from oestrogen and progesterone metabolites in urine (Wilcox et al. 1995). Functionally-based studies of the relationship between the cycle and preference or behaviour changes must choose some way to identify ovulation and this conceptive phase, and often base this on the hormonal or physiological changes of the cycle.

There is a complex hormonal interplay affecting ovulation and the menstrual cycle (details in e.g. Sherwood 1989). Gonadotrophic-releasing hormone from the hypothalamus causes release of follicle-stimulating hormone from the pituitary gland. This leads to the maturation of a handful of follicles in the ovaries, and stimulates the ovaries to produce oestrogen. This oestrogen increase supports the development of an ovum, thickens the lining of the uterus, and thins cervical mucus. Next, follicle-stimulating hormone levels dip, and then rise again as luteinising hormone release from the pituitary gland peaks sharply, triggering ovulation and marking the end of the follicular phase and the beginning of the luteal phase. The empty follicle now becomes a progesterone-producing corpus luteum. This progesterone thickens the mucus at the cervix and supports the development of the womb lining. So long as no fertilised ovum is implanted, progesterone levels eventually drop, and prostaglandins released from the lining of the uterus precipitate uterine contraction and menstrual bleeding.

A key methodological challenge for psychological research on the effects of the cycle is to choose an optimal method to determine ovulation and the conceptive phase. There are a number of possibilities, each with strengths and weaknesses. Roughly in order of accuracy of detecting ovulation, clinical methods include pelvic ultrasonography (Renaud et al. 1980; Garcia et al. 1981; Wetzels et al. 1982; Vermesh et al. 1987; Behre et al. 2000), which can trace the increase in size of the follicle and its apparent disappearance as the egg is released at ovulation; measurement of the levels of the hormones estradiol, progesterone or luteinizing hormone from blood plasma (Vermesh et al. 1987; Behre et al. 2000); or measurement of luteinizing hormone from urine (Guida et al. 1999). These necessitate inconvenient and costly daily attendance of the subject at a clinic, although this last method can equally be carried out by fairly accurate commercially available kits (Vermesh et al. 1987; Gudgeon et al. 1990; Miller & Soules 1996; Behre et al. 2000). Methods such as measurement of salivary ferning, inspection of cervical mucus, basal body temperature rises, and the noting of mittelschmertz (a sensation or pain associated with the release of the ovum from the follicle) tend to be fairly unreliable indicators of ovulation (Kerin 1982; Wetzels et al. 1982; Vermesh et al. 1987; Guida et al. 1999) and require a fair degree of participant compliance, commitment and training.

Non-clinical methods may also be used. Although the menstrual cycle can vary in length both between and within women (Burley 1979), ovulation occurs fairly reliably 14 days prior to the onset of menses (Fluhmann 1957; Lein 1979). Ovulation can therefore either be predicted from information on a woman's average cycle length and the date of onset of last menses, or determined retrospectively from information on the woman's next onset of menses. These types of 'counting' methods are conventionally used in the behavioural literature. Their advantages include that, relative to the clinical methods, they are less invasive and more convenient, and thus often more acceptable both to ethics committees and to female participants. They can be used outside of the laboratory or remotely, such as in internet-based data collection. Less obviously, researchers can conceal from the participants that they are investigating cycle effects, thereby reducing the likelihood of demand characters. However, the counting method also has some disadvantages. Women may forget or be careless in reporting menses dates. The onset of menses can be hard to identify for those women who experience a particularly light flow at onset (Walker 1997). Perhaps more seriously, menstruation does not confirm that ovulation has occurred, and in fact anovulatory cycles are not uncommon, especially in the early postmenarcheal phase which constitutes the first three years after menarche (Apter 1980; Apter & Vihko 1983; Ibanez et al. 1999). Reports of the prevalence of anovulatory cycles amongst adult women range from less than 1 % (Behre et al. 2000), to 10 % (Morrell et al. 2002), to between 10 % in sedentary women to around 50 % of regularly exercising women (De Souza et al. 1998). Although anovulatory cycles share some of the hormonal characteristics of ovulatory cycles, it is not clear *prima facie* that we should expect them to provoke an identical profile of behavioural or cognitive changes.

Despite methodological problems such as these, psychological research, which tends to make use of counting methods to determine cycle phase, has presented fairly consistent findings. One branch of this research has examined female attractiveness, which seems to be rated more highly at the high-fertility compared with the low-fertility phase of the cycle. Attractiveness changes across the cycle have been documented in a wide range of specific modalities; studies are reviewed in more detail in Chapter 4. A typical experimental design asks raters to compare high- and low-fertility recordings of a set of females, and choose the more attractive. Recordings might be a photograph, or a body odour sample, for instance. Cycle-related changes in attractiveness are reported to be subtle (e.g. Roberts et al. 2004). However, women do not tend to be perceived in single modalities in the real world, and changes may be more easily detectable when suites of modalities are presented simultaneously rather than in isolation (c.f. Møller & Pomiankowski 1993; Rowe 1999; Candolin 2003). To understand how cycle changes in attractiveness influence others' perceptions, it may therefore be necessary to present stimuli in more than single modalities, and this question is taken up in Chapter 4. Indeed, a concern relating to this interplay between single and multiple modalities has recently been mooted in a related research area. Psychological research often makes use of single-modality recordings, such as facial photographs, to evaluate or represent the attractiveness of an individual. Yet the validity of using such single-modality traits in this way has been put into question (Rubenstein 2005; Lander 2008). Chapter 4 reports two complementary studies, one of which sought to determine the relationship between single-modality traits and overall attractiveness, and the second of which investigated how changes in attractiveness across the menstrual cycle in single traits may be relevant to more holistic contexts. The studies find that between-individual differences in visual and vocal attractiveness are related and contribute to ratings of individual attractiveness, but that within-individual changes in attractiveness across the cycle do not correspond highly across modalities.

The question of whether changes in a woman's attractiveness function to her benefit, or to the benefit of the perceiver, is discussed later. It is perhaps not surprising to find though that female attractiveness changes do not function in isolation from female cognitive changes. Again, changes in woman's mate preferences across the cycle are well-documented, and are reviewed in Chapter 5. In essence, when women make mate choice judgments, they appear to be most attuned to markers of genetic quality, and to be averse to cues of resemblance, during the conceptive phase of the cycle (Jones et al. 2008a). Yet an individual who is desirable with respect to one of these qualities may be undesirable because of the other; the question of how women might weight these two dimensions has so far gone unexamined (Roberts & Little 2008). Chapter 5 reports two studies which investigate this, the first a larger single-shot preference study, and the second a two-session study with relation to the menstrual cycle. The studies find that judgments of self-resemblance and sexual dimorphism may be described by a context-dependent, heirarchical rule; the study reported presents no evidence for menstrual cycle effects on these judgments.

Chapter 4: Single- and multi-trait attractiveness

4.0.1 Components of attractiveness

Although judgments of physical attractiveness can be made with reference to a wide range of physical attributes, most work on human attraction has employed single-modality stimuli such as photographs of faces or recordings of voices, with ratings of these traits often assumed to be a proxy for the overall attractiveness of an individual (review in Rhodes 2006). Research has shown that both face and voice (Zuckerman et al. 1991), face and body (Mueser et al. 1984; Brown et al. 1986; Raines et al. 1990b; Peters et al. 2007), and face and dynamic expressiveness (Riggio et al. 1991) can contribute individually to perceptions of individual attractiveness. Yet work on the contributions of various modalities (in particular, the relative importance of face and voice) to social impression formation (e.g. perceptions of dominance or trustworthiness at zero acquaintance) has emphasised that the contribution of each modality is dependent both upon the context in which judgments are made and the context in which stimuli are recorded (Ekman et al. 1980; Zebrowitz-McArthur & Montepare 1989; Raines et al. 1990a, b). To understand how the different modalities contribute to perceptions of attractiveness at zero acquaintance, then, research specific to judgments of attractiveness in zero acquaintance social contexts is necessary. Such research has shown that the face is more important than the rest of the body to judgments of static whole-body attractiveness (Peters et al. 2007), and that facial attractiveness and dynamic expressiveness are more important than attractiveness of body or dress (Riggio et al. 1991). Yet an examination of the various contributors to overall personal attractiveness in more naturalistic conditions, and in particular a consideration of the important modulatory effect of vocal attractiveness to overall attractiveness (Zuckerman et al. 1991), has so far been lacking.

Related to the question of the contributions of the different modalities to attractiveness judgments, there is a controversy over whether different traits each provide similar information on an individual's attractiveness. Correlations between attractiveness ratings are predicted on the theoretical footing that physical traits may provide concordant information about an individual's suitability as a partner (Møller & Pomiankowski 1993). In line with this, ratings of faces and voices (Feinberg et al. 2005a; Saxton et al. 2006), faces and odour (Rikowski & Grammer 1999), and male personality and odour (Havlicek et al. 2005) are positively correlated within an individual. Yet other findings have been mixed. Research has found a relationship in women but not in men between the ratings of faces and bodies (Thornhill & Grammer 1999; Peters et al. 2007), and between the ratings of faces and facial movement (Morrison et al. 2007). Reports of the correspondence between attractiveness ratings of facial photographs and silent videos of the moving faces have reported variously that they correspond in stimuli of both sexes (Brown et al. 1986; Roberts et al. in review), in female but not male stimuli (Lander 2008; Penton-Voak & Chang 2008), and not in female stimuli (Rubenstein 2005). Similarly, correlations between attractiveness ratings of voices and video recordings of the faces have been reported as lacking (Raines et al. 1990b) or as present only from opposite-sex judgments (Lander 2008). Finally, no significant correlations have been found between attractiveness ratings of voices and facial or physical appearance (see Oguchi & Kikuchi 1997; Lander 2008), between face, body and dynamic movement (Riggio et al. 1991), and between ratings from photographs and ratings by peers (Kniffin & Wilson 2004). Such inconsistencies cast doubt on the validity of using single measures such as static facial photographs to serve as proxies for overall individual attractiveness.

Experiment 4a, then, explicitly tests the contribution of, and relationship between, specific components in different modalities (i.e. face, clothed body, and speech) to overall judgments of attractiveness. We first video recorded a set of target individuals as they introduced themselves. We then asked one group of raters to rate each of these introductions for attractiveness, and a second, independent group of raters to rate the various components (face, clothed body, speech) of these videos for attractiveness. We investigated the relative importance of each of the isolated components to the overall rating, and whether there were correlations between the individual components.

4.0.2 Attractiveness changes across the cycle: a multi-modal signal?

Differences in attractiveness between individuals are apparent from our everyday experience, but attractiveness can also vary within individuals. There has been a good deal of recent research interest in how a woman's attractiveness varies across the menstrual cycle (for a review, see Gangestad & Thornhill 2008). Female conception risk varies across the cycle, being highest in the days leading up to and including ovulation (Wilcox et al. 1995). During this fertile window, ratings of female attractiveness have been found to increase. Several studies have reported that female bodily odours (armpit and genital) are rated as more pleasant at the high- compared with the low-fertility phase of the cycle (Doty et al. 1975; Poran 1994; Singh & Bronstad 2001; Kuukasjarvi et al. 2004; Havlicek et al. 2006). In normally-cycling women, facial photographs, whether masked to conceal the hairstyle or with the hairstyle visible, are rated as more attractive at peak fertility (Roberts et al. 2004), as are recordings of women's voices (Pipitone & Gallup 2008), and men judged whole-body photographs of women to be more attractive at the fertile compared with non-fertile phase (Schwarz & Hassebrauck 2008). Women at peak fertility are judged as "trying to look more attractive" in their clothing choices (Haselton et al. 2007), and higher levels of estradiol, a hormone associated with ovulation, have been linked to the wearing of more revealing clothing (Grammer et al. 2004). Although the majority of studies have focused on laboratory-based ratings of the attractiveness of single traits such as odour or face, recent research has also found that financial earnings of female lap-dancers increase at high-fertility (Miller et al. 2007). The reason for the increased earnings is unknown, but if it is attractiveness based, this suggests that attractiveness changes may have measurable effects in the real world, at least in explicitly sexual contexts. However, it is not yet known whether changes in attractiveness are apparent when raters are presented with a suite of traits such as would be experienced in the real world in more social contexts, and whether these changes can be detected simultaneously in a variety of characters.

One possibility is that changes in attractiveness across the menstrual cycle are particularly evident in some traits or modalities and less apparent in others. It has been suggested that attractiveness changes may have behavioural impacts within established sexual relationships (Havlicek et al. 2006); signalling theory predicts that these sorts of collaborative relationships may lead to the evolution of particularly subtle signals that are most evident to a close partner (Krebs & Dawkins 1984; and Provost et al. 2008 make this argument specifically in relation to fertility). In addition, there are likely to be morphological constraints on attractiveness changes. For instance, cyclic changes have been demonstrated in the symmetry of soft tissue such as that making up the ears, digits, and breasts (Manning et al. 1996; Scutt & Manning 1996), and so traits in which soft tissue changes are most evident may be particularly prone to provoke differences in judgments of attractiveness.

Alternatively, it may be that changes in attractiveness associated with the high-fertility phase function as a single 'ornament' that is evident across a range of traits. If each trait in one individual presents multiple concordant information, this can lead to increased detection, recognition, discrimination, and memorability of the underlying information as well as constituting back-up signals that ensure that the message reaches the recipient (Møller & Pomiankowski 1993; Rowe 1999; Candolin 2003). As such, previous reports that changes in rated attractiveness between high- and low-fertility phases are very subtle (e.g. Roberts et al. 2004) could underestimate the size of the effect. If physical attractiveness increases in a number of traits simultaneously, this may lead to changes in attractiveness that are more readily perceptible when judgments are made with reference to multiple traits.

Accordingly, Experiment 4b set out to test the effect of, and concordance between, different modalities in perceptions of female attractiveness across the menstrual cycle by recording women at high- and low-fertility phases, contrasting normally-cycling women with a control group of users of hormonal contraception. Women were recorded introducing themselves as they might when they met someone for the first time, a procedure chosen in order to establish ecological validity in the context of human mate choice. One group of raters was asked to judge which of the high and low fertility video recordings was more attractive, and a separate group made the same judgments in relation to component stimuli (e.g. facial photograph, voice recording).

4.1 Experiment 4a: Components of attractiveness

4.1.1 Methods

Fifty individuals (aged 18 – 39; mean ± SD 24 ± 4 years; 25 males) were recruited from the student population and from social contacts. Individuals provided informed consent to participate in a study on the contribution of different components to impression formation including initial attractiveness judgments. Each individual was asked to introduce themselves as though meeting someone new, and to speak for at least 20 s. Sitters were recorded at a distance of approximately 1.5 m from the video camera (Sony DVD DCR200E) while seated in front of a white backdrop. The zoom was adjusted manually to frame the participant from top of head to mid-thigh. White balance, exposure and focus were manually set and held constant. Sitters were asked to introduce themselves as they might upon meeting someone for the first time in a pub or bar. They were prompted to talk on any subject they wished, including their studies, hobbies, work, free time, holidays or weekend activities. Sitters were free to request that the researcher leave the room for the duration of the recording, and to have a repeat take of the recording, for example if they failed to speak for 20 s. Sitters were photographed (Canon Powershot) in front of the white backdrop as they adopted a neutral expression looking straight at the camera. Seated photographs of the face alone, and then standing photographs of the entire face and body with arms held vertical and parallel with their sides, were taken. Individuals were also recorded (M-Audio Microtrack 24/96, Audio-Technica ATR55 Telemike Shotgun cardioid condenser microphone) as they read out a semantically neutral sentence: "The quick brown fox jumped over the lazy dog". All recordings were made in the same location within a windowless recording room lit with standard fluorescent lighting.

Facial photographs were cropped with an oval around the face outline, concealing the majority of the hair (Corel Paint Shop Pro Photo X2) and normalised in size with reference to interpupillary distance (c.f. Burriss & Little 2006), and body photographs were cropped from neck to fingertips (i.e. omitting the head) and normalised for height. Size normalisation was carried out with specialist image software (Psychomorph; Tiddeman et al. 2001). Resultant image dimensions were 1276 x 1276 pixels (face) and 1181 x 1259 pixels (body). Video recordings were edited to a duration of 20 s, cropped to dimensions of 400 x 480 square pixels and encoded as 25 f.p.s. QuickTime movies using the MPEG-4 codec (Adobe After Effects 7.0). Soundtracks were normalised for amplitude to ensure volume similarity between recordings (Adobe Audition 2.0).

Raters (aged 18 – 32; mean ± SD = 20 ± 3 years) were recruited from the student population of a different university from that at which sitters were recruited, and participated in one of three separate sessions in exchange for course credit. Participants filled out basic demographic information on an anonymous, paper-based questionnaire. In two of the sessions, raters (N = 26; 7 males) were presented with the complete 20 s video of each self-introduction. Raters reported themselves to be white European except for one who did not answer; one rater reported herself to be homosexual and the rest reported themselves to be heterosexual. In a third session, raters (n = 13; 3 males) were presented with masked neutral face photographs, upper body images, and finally video soundtracks. Small numbers of raters are generally sufficient for judgments of attractiveness because attractiveness judgments are highly homogeneous between raters (e.g. Feinberg et al. 2005a; Jones et al. 2005b). Raters from the third session also rated a subset (n = 39) of the voice recordings of the neutral sentence. Female stimuli were presented before male stimuli in each case. Raters reported themselves to be white European except for one who described herself as Asian; one rater reported herself to be bisexual and the rest reported themselves to be heterosexual. Sessions took place in a classroom where stimuli were projected onto a screen and sounds played over a speaker system. Raters rated each stimulus for attractiveness on a scale from 1 – 7 (anchors "unattractive" and "very attractive") on a sheet of paper. They were told that the sitters were recruited from a normal student population, and with reference to this they should endeavour to use the whole rating scale if possible. They were instructed not to respond in any way which could transmit to others in the room their reaction. The mean attractiveness rating of each component for each sitter was calculated.

Analysis was carried out using SPSS 15.0. Kendall's coefficient of concordance was used to examine concordance; one rater was confused at the beginning of the session and did not rate the first 25 masked faces, and so her ratings were excluded from the concordance calculations for all of the masked faces. If raters omitted to rate a stimulus (n = 55 of a possible 3225 ratings) values were replaced by the mean value given to that stimulus so that the rest of the rater's ratings could be included in the analysis. Kendall's tau is used for correlational analysis because some sets of ratings were non-normally distributed once ratings had been split for separate analysis for male and female stimuli. All assumptions of regression were satisfied, including that of no multicollinearity, as measured with reference to correlation between pairs of variables, VIF, tolerance, loadings and determinant of the correlation matrix (Field 2005). Ratings of the neutral sentence were not included in the regression analysis because they correlated highly with ratings of the soundtrack (r = .660, $p < .001$). The enter method of regression was used because no assumption was made as to the relative importance of each component to overall attractiveness judgments (Field 2005). Mean ratings of the overall video was unit of analysis, and mean ratings of each component as the dependent variables, in the regression, because this captured the logic of investigating how the ratings of the components relate to that of the overall rating.

4.1.2 Results

Agreement between judges was significant (face: W = .480; body: W = .441; soundtrack: W = .339; all $p < .001$). Linear regression

	all stimuli		*male stimuli*		*female stimuli*	
	r	*p*	*r*	*p*	*r*	*p*
face	.426	< .001	.350	. 015	.463	.001
body	.460	< .001	.425	. 004	.374	.010
soundtrack	.320	.001	.367	.012	.294	. 044

Table 1: Correlations of attractiveness ratings between individual components and rating of complete video

was used to analyse the relationship between the mean ratings of the complete video and the various components (dependent variable: mean attractiveness score for overall video; independent variables: mean attractiveness scores for each of the three stimuli types). The overall model was significant (adjusted r^2 = .508, F = 17.85, df = 46, p < .001). Mean ratings of the face (β = .349, t = 3.07, p = .004), body (β = .343, t = 2.89, p = .006) and soundtrack (β = .278, t = 2.60, p = .013) were independently and positively related to mean ratings of the complete video.

Even though ratings were made by separate groups, there were significant correlations between the mean rating of the complete video and the mean ratings of the various components, both for all stimuli or for male and female stimuli separately (Table 1). Correction for multiple comparisons is not made because the correlations are not entirely independent of each other (i.e. if there is a correlation between face and video when all stimuli are judged, we would anticipate a similar relationship when the stimuli are restricted to the men or the women) and so a straightforward correction is not possible. Mean ratings of the various components were also positively and significantly correlated with each other (face/body: $\tau = .350$, $p < .001$; face/soundtrack: $\tau = .196$, $p = .050$; body/soundtrack: $\tau = .197$, $p = .049$).

4.1.3 Discussion

The experiment builds upon previous research which has examined how different components such as the face and body combine to form a more holistic impression of personal attractiveness (e.g. Riggio et al. 1991; Peters et al. 2007). It also sought to investigate conflicting findings showing that trait attractiveness in different modalities is (e.g. Brown et al. 1986; Rikowski & Grammer 1999; Thornhill & Grammer 1999; Feinberg et al. 2005a; Havlicek et al. 2005; Saxton et al. 2006; Feinberg et al.; Roberts et al. in review) or is not (e.g. Raines et al. 1990b; Riggio et al. 1991; Oguchi & Kikuchi 1997; Rubenstein 2005; Lander 2008) positively correlated. It set out to categorise the relative importance of face, body and speech to overall judgments of attractiveness, and to investigate the relationships between attractiveness judgments of face, body and speech. A set of target individuals introduced themselves on video tape. The self-introduction video was designed to correspond to zero acquaintance social contexts. One group of raters rated each of these introductions for attractiveness. A second, independent group of raters rated the various components (a facial photograph of a neutral expression, masked to conceal the hair; the upper body, revealing clothing choice; and the soundtrack of the video) for attractiveness.

Although one group of individuals rated the full videos, and a second group rated the face, body and speech, the mean ratings of the face, clothed body and speech (in order of relative magnitude) were all significant predictors of the mean ratings of the complete videos. Ratings of the individual components (face, clothed body, speech) also correlated significantly with each other. The correlation between the face and body demonstrated a medium-sized effect, while correlations with the speech were of a small effect size.

Our results are not wholly consistent with previous research showing little relationship between attractiveness judged from a voice recording compared with attractiveness judged from a video recording (Raines et al. 1990b, using videos complete with sound; also Lander 2008 in same-sex judgments of silent videos), and with findings of low or no relationship between the attractiveness ratings of facial photographs and the attractiveness ratings of silent video recordings of the speaking faces (Rubenstein 2005; Lander 2008; and Penton-Voak & Chang 2008 in recordings of males), but there are a number of methodological differences. Unlike the present study, vocal cues were only available to the raters of the videos in one (Raines et al. 1990b) of these previous studies. Vocal cues can modulate judgments of attractiveness (Zuckerman et al. 1991; Miyake & Zuckerman 1993; Zuckerman et al. 1993). Previous studies also used shorter video extracts (between two and 10 s compared to the 20 s of the current study), and in all but one (Penton-Voak & Chang 2008) of the studies attempted to eliminate (Rubenstein 2005; Lander 2008) or pre-specify (Raines et al. 1990b) emotional information and personal expression, qualities that affect perceptions of attractiveness (Raines et al. 1990a, b; Penton-Voak & Chang 2008). Exaggerated female facial movements are associated with enhanced female attractiveness (Morrison et al 2007; Penton-Voak & Chang 2008), and it is possible that dynamic emotional or sexually dimorphic expression strengthens the relationship between static and dynamic attractiveness. In the present study, the inclusion of vocal information in longer and more naturalistic recordings where personal and emotional expression are allowed represents most realistically how an individual appears in the real world than do emotion-free or soundless dynamic facial images. Although the current study makes use of videos that are still a step removed from the real world, the finding that individual components such as static facial images relate to overall attractiveness suggests that ratings of individual components may yet be valid indices of overall individual attractiveness, and supports such usage.

In conclusion, the results suggest that studies that employ visual stimuli alone to assess individual attractiveness in a mate choice context may be prone to a systematic source of experimental noise in that they do not consider the role of vocal attractiveness. Assessment of both vocal and visual components (c.f. Zuckerman et al. 1991; Miyake & Zuckerman 1993) would assist in a fuller understanding of an individual's attractiveness. Nevertheless, in replicating results showing that the face is more important than the body (Peters et al. 2007), and in extending this to show that facial attractiveness also has greater relative import than speech attractiveness, the findings support the prevalent use of facial photographs as proxy measures of an individual's attractiveness within human attraction and mate choice research. Future research might consider the individual differences and the contextual effects that may moderate this relationship.

4.2 Experiment 4b: Attractiveness and the menstrual cycle

Following on from these findings of Experiment 4a of the relevant importance of these different components to overall judgments of attractiveness, Experiment 4b set out to discover how different components might affect judgments of changes in attractiveness across the menstrual cycle.

4.2.1 Methods

4.2.1 a Female targets

Eighty-eight women were recruited from amongst university students and social contacts for a within-subjects study on the stability of perceptions of attraction and attractiveness and consented to have recordings rated by participants outside the university; they were not told the study hypotheses. Before they attended any study session, women completed an initial demographic questionnaire which included questions on age, relationship status, ethnicity, hormonal contraceptive usage, date of last and next predicted menses, and details of average cycle length. Non-users of hormonal contraceptives all reported a usual cycle length of between 25 and 35 days. Users of hormonal contraceptives were assigned to the CONTRACEPTIVE group, and non-users were assigned to the NON-CONTRACEPTIVE group.

4.2.1 b Session scheduling

Sessions were scheduled to occur at high- and low-fertility phases, estimated from data provided in the initial questionnaire. While the CONTRACEPTIVE group women do not differ in conception risk across the cycle, the same terminology is used here for purposes of comparison. The high-fertility phase constituted recordings made 14 – 19 days before next menses (i.e. Day 10 – 15 in a 28-day cycle, where Day 1 is the day of menses onset) and the low-fertility phase constituted recordings made 2 – 12 days before next menses (i.e. Day 17 – 27 in a 28-day cycle) (Wilcox et al. 1995). Because inter-individual variation in cycle length is mainly accounted for by differences before ovulation, the backwards-counting method of ascertaining fertility status is more accurate than forward-counting from previous menses (Johnston et al. 2001). Women were contacted by email to seek information regarding onset of next menses subsequent to the sessions, and NON-CONTRACEPTIVE group women returned for a third (n = 2) or fourth session (n = 1) as necessary to obtain high- and low-fertility recordings from all participants. Four of the CONTRACEPTIVE group women also returned for third sessions due to technical problems with the recording equipment. We did not exclude women who attended a recording session just prior to menstruation because research demonstrates large individual variability in the link between cycle phase and mood (Walker 1997); in any case, only one woman from the NON-CONTRACEPTIVE group final sample (see below) attended a recording session in the three days prior to menstruation. Five women used in the final sample (see below) did not experience regular menstruation. Of these five, two women took oral contraceptives with a 28-day cycle and so the first day of the pill packet was defined as 'Day 1'. The remaining three women either used contraceptive implants (n =1) or oral contraceptives with no 'Day 1' (n = 2). We scheduled the sessions of one of these women eight days apart, and allocated her first session to the 'high-fertility' phase and her second session to the 'low-fertility' phase. We scheduled the sessions of the other two women 14 days apart, and allocated their first session to the 'low-fertility' phase and their second session to the 'high-fertility' phase.

4.2.1 c Recording procedure

Women removed all make-up using provided make-up remover at least five minutes prior to the recordings to allow the subsidence of any skin flushing due to the removal process. Four recordings were made: 1) a direct frontal facial photograph in which women were instructed to adopt a neutral expression as they might do for a passport photograph; 2) a facial photograph in which women were instructed to adopt a natural or relaxed expression as they might do if photographed by a friend; 3) a photograph of the woman's full body as she stood facing the camera with arms parallel with the sides; 4) a video recording where women introduced themselves, as they might do if they met someone for the first time, speaking for at least 20 seconds while seated at a distance of 1.5 m from the video camera. Pilot testing indicated that some subjects had difficulties with this last task, and so all subjects were given a list of possible subjects (work, study, home life, weekend or evening activities, hobbies, recent holidays) as ideas for content. With the aim of maintaining a degree of anonymity, subjects were told not to reveal their name. At the second session, women were told that even though they had carried out the video recording previously, they were to approach it as though it were a completely new task. All recordings were carried out by one of the female researchers (see Acknowledgments). Researchers were distributed at random between high- and low-fertility recordings and CONTRACEPTIVE and NON- CONTRACEPTIVE groups.

Upon completion of the study, women filled in another questionnaire to confirm answers from the first questionnaire. Women were also interviewed by one of the female researchers (see Acknowledgments) about their understanding of the study hypotheses. They were given the opportunity to talk generally, and were also specifically asked why they thought they were recorded on two separate sessions. Of the 34 women used in the final sample (see below), only three linked the recording sessions to the menstrual cycle. Two women (from the NON-CONTRACEPTIVE group) assumed that one of the study aims was to investigate the effect of the menstrual cycle on attractiveness, and the third (from the CONTRACEPTIVE group) suggested that we may have been interested in postural changes across the menstrual cycle. Frequent suggestions from the remaining women were that the reason for obtaining two sets of recordings was to control for the effects of practice or to obtain sets of stimuli for use in future psychological experiments.

4.2.1 d Final sample

Of the 88 recruits (54 allocated to the NON-CONTRACEPTIVE group), 18 women did not attend two sessions; 17 did not attend in both high- and low-fertility sessions; one chose not to remove make-up; five chose not to provide video or sound recordings; and the video file of one woman became corrupted and could not be used. These exclusions reduced the sample pool to 16 normally cycling women (age 17 – 32; M = 24.7, SD = 4.4 yrs; *n* = 11 in a relationship; *n* = 11 white or mixed white, European and *n* = 5 Chinese; eight of whom attended the first recording session during the high-fertility phase) and 30 users of hormonal contraceptives. A random selection of 18 of the 30 women from the CONTRACEPTIVE group (age 18 – 31; M = 23.1, SD = 4.1 yrs; *n* = 10 in a relationship; *n* = 12 white or mixed white, European and *n* = 1 Hispanic; *n* = nine of whom attended the first recording session during the high-fertility phase) were prepared as stimuli. The reason for this discrepancy in sample numbers between the CONTRACEPTIVE and NON-CONTRACEPTIVE groups is that some of the women were not able to confirm onset of next menses until after the rating session had taken place, and ratings of recordings that transpired not to have been made in the correct phase had to be discarded.

4.2.1 e Stimuli

Neutral-expression facial photographs were cropped (Corel Paint Shop Pro Photo X2) and masked with an oval around the face outline, concealing the majority of the hair (Jones et al. 2004a); this stimulus type will henceforth be referred to as the "MASKED FACE". Relaxed-expression facial photographs were cropped from neck to top of hair ("FULL FACE"). Both kinds of facial photograph were normalized in size with reference to the interpupilary distance (Burriss & Little 2006), resulting in image dimensions of 1276 x 1276 pixels (MASKED FACE) or 1181 x 1259 pixels (FULL FACE). The photographs of the upper body were cropped from neck to fingertips and normalized for height ("TORSO"), resulting in images of 1181 x 1259 pixels. Image size normalization and all masking was carried out using specialist software (Psychomorph; Tiddeman et al. 2001). Video recordings ("FULL VIDEO") were edited to a duration of 15 s, starting from the first words spoken, cropped to dimensions of 400 x 480 square pixels, and encoded as 25 f.p.s. QuickTime movies (Apple Inc.) using the MPEG-4 codec (Adobe After Effects 7.0). The soundtrack was normalized for amplitude to ensure volume similarity between recordings (Adobe Audition 2.0). Silent versions of each video ("SILENT VIDEO"), and separate soundtracks for each video ("SOUNDTRACK"), were also created (Adobe Audition 2.0).

4.2.1 f Raters

Raters were recruited from students of a different university to minimize potential familiarity with the stimuli. Students were recruited principally from amongst psychology undergraduates, and received course accreditation for participation. These raters participated in one of five separate sessions. Raters were presented with pairs of stimuli. Each pair comprised stimuli recorded from the same woman, with one stimulus recorded during the high-fertility and one during the low-fertility phase. For each pair, raters indicated which stimulus they judged more attractive. Raters in sessions 1 (n = 14, 4 male) and 3 (n = 15, 7 male) were presented with all stimulus types except for the FULL VIDEO. Raters in sessions 2 (n = 9, all female) and 4 (n = 17, 7 male) were presented only with the FULL VIDEO. Stimuli presentation side (photographic stimuli) or order (sound/video stimuli) was reversed between sessions, so that stimuli were presented in reverse order in session 3 compared to session 1, and session 4 compared to session 2. Stimulus type presentation order was also reversed between sessions 1 and 3 (i.e. so that raters in session 1 saw the MASKED FACES first, and raters in session 3 saw them last). Sessions took place in a classroom where photographs were projected side-by-side, or videos presented sequentially, onto a screen placed approximately between 2 and 5 m from raters, and sounds played sequentially over a speaker system. Raters indicated their choices on a sheet of paper; they were instructed at the beginning of the session not to confer, or to react in any way which could transmit to others in the room how they were likely to respond. After the session, raters completed a questionnaire where they were asked to provide data including age, sexual orientation, sex, and ethnicity. They were reminded that they did not have to answer any question that they were not comfortable with, and that all data would be stored securely and anonymously. Raters from sessions 1 and 3 were aged 18-23 (mean ± SD = 19.0 ± 1.2); described themselves as heterosexual (with the exception of one homosexual and one non-respondee); and gave their ethnicity as white European (with the exception of two mixed race British, two Chinese, and one non-respondee). Raters from sessions 2 and 4 were aged 18-32 (mean ± SD = 20.0 ± 3.1); described themselves as heterosexual (with the exception of one homosexual); and gave their ethnicity as white European (with the exception of one non-respondee).

4.2.1 g Statistical analysis

For each woman, the proportion of times that her high-fertility recording was chosen as more attractive was calculated separately for each of the six stimulus types (FULL VIDEO, MASKED FACE, FULL FACE, TORSO, SOUNDTRACK, SILENT VIDEO). The denominator was adjusted accordingly if ratings were omitted. Prior to analysis, 0.5 was subtracted from the proportions to enable use of the test of the intercept (which tests whether the grand mean differs significantly from zero) to determine whether the high-fertility phase recordings were chosen significantly more often than chance (c.f. Haselton et al. 2007). ANOVA was carried out in SPSS 15.0. Some of the sets of data were non-normally distributed (Shapiro-Wilk, $p < .05$), but ANOVA is robust to violations of the assumption of normal distribution of data (Subrahmaniam et al. 1975; Field 2005). Data fulfilled other assumptions of the statistical tests. To answer the question of whether high fertility was more detectable in some modalities than others, analysis was carried out both with female target as unit (to control for individual differences in magnitude of change between high- and low-fertiliy), and also with rater as unit of analysis (to control for individual differences in rater ability to select the high-fertility recording as more attractive). The first analysis allowed simultaneous investigation of the effects of contraceptive usage, while the second allowed simultaneous investigation of the effects of rater sex.

4.2.2 Results

4.2.2 a Analysis by female target

Mixed model analysis (within-subject factors: six stimuli types; between-subject factors: CONTRACEPTIVE or NON-CONTRACEPTIVE group; whether high- or low-fertility recording made first) revealed that overall, the high-fertility recordings of women from the NON-CONTRACEPTIVE group were selected significantly more often than those of the women from the CONTRACEPTIVE group ($F_{1,30} = 5.924$, $p = .021$), with a medium effect size ($r = 0.41$). Women were selected as more attractive in the high-fertility session if that coincided with the first recording session ($F_{1,30} = 10.31$, $p = .003$), and there were significant differences in the selection rates of different stimuli types ($F_{5,150} = 2.28$, $p = .049$). Inclusion of all stimulus types in the same analysis assumes that raters in sessions 1 and 3 were acting identically to raters in sessions 2 and 4, and so a confirmatory analysis was also carried out with the exclusion of the FULL VIDEO ratings. Results were qualitatively identical. Finally, less powerful univariate analysis of each modality separately showed in each case a non-significant tendency for recordings to be selected at greater rates from amongst women in the NON-CONTRACEPTIVE compared to the CONTRACEPTIVE group (all $p > .07$).

Following on from the significant effect of group (CONTRACEPTIVE or NON-CONTRACEPTIVE) on high-fertility selection proportions, we carried out a subsequent mixed-model analysis for the two groups separately (within-subject factor: six stimuli types; between-subject factor: whether high- or low-fertility recording made first) which showed that stimulus recordings from the high-fertility phase of the NON-CONTRACEPTIVE group but not the CONTRACEPTIVE group were chosen significantly more often than chance (test of the intercept: NON-CONTRACEPTIVE group: $F_{1,14} = 5.914$, $p = .029$; CONTRACEPTIVE group: $F_{1,16} = .266$, $p = .613$). Amongst the NON-CONTRACEPTIVE group, there was no significant effect of stimulus type on the proportions of high-fertility recordings selected ($F_{5,70} = .77$, $p = .574$), but a greater proportion of high-fertility recordings were chosen as more attractive if the high-fertility phase coincided with the first session ($F_{1,14} = 6.55$, $p = .023$). Amongst the CONTRACEPTIVE group, there was no significant effect of stimulus type on the proportions of high-fertility recordings selected ($F_{5,80} = 1.99$, $p = .089$), and no significant differences between the selection rates of the first compared to the second session ($F_{1,16} = 3.27$, $p = .090$). Again, results are qualitatively identical if the analysis is restricted to the five stimulus types judged by raters in sessions 1 and 3 only. Results are represented in Figure 5.

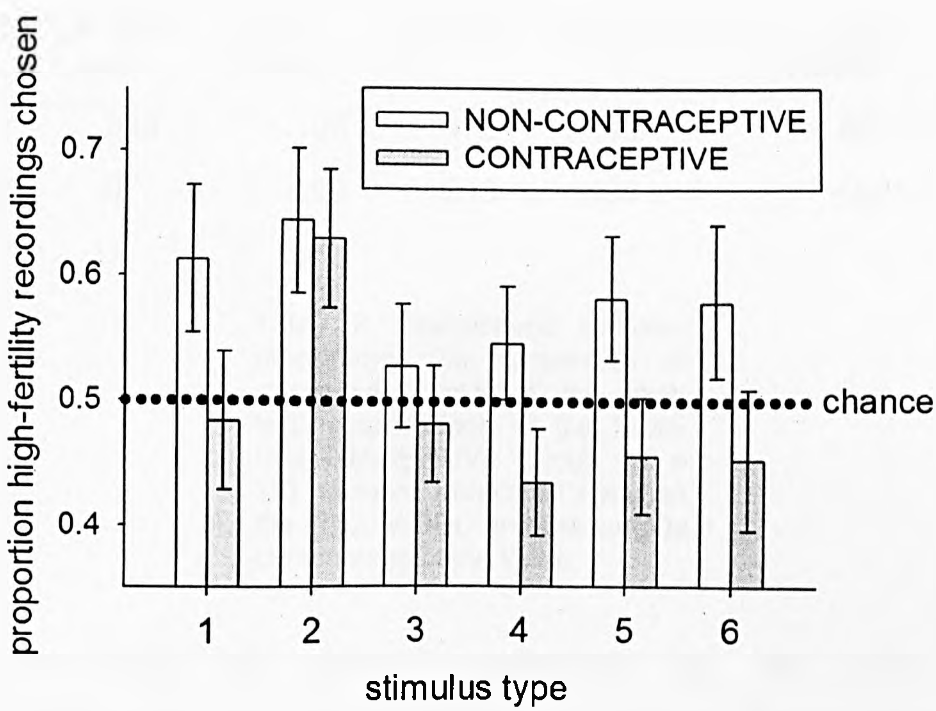

Figure 5. Estimated marginal means ± SE, proportion of high-fertility recordings chosen (female as unit of analysis), controlling for order of recording sessions. Stimulus type: 1: MASKED FACE, 2: FULL FACE, 3: TORSO, 4: SOUNDTRACK, 5: SILENT VIDEO, 6: FULL VIDEO

Kendall's coefficient of concordance was used to investigate whether the NON-CONTRACEPTIVE women who were more attractive at high-fertility in one modality (as indexed by the proportion of raters selecting the high-fertility recording) were also more attractive at high-fertility in another modality. There was no evidence of concordance between modalities within individual women (W = .036, p = .720; female as unit of analysis; proportion across six modalities). That is, if a woman was selected by a large number of raters as more attractive at high fertility in one modality, this would not lead us to predict that she would be so selected in another modality. The specific correlations between the overall

	MASKED FACE	FULL FACE	TORSO	SOUNDTRACK	SILENT VIDEO
τ	.335	.196	.472	.149	.661
p	.077	.298	.013	.438	<.001

Table 2. Correlations between proportions (i.e. proportion of raters who selected the high-fertility recordings of the NON-CONTRACEPTIVE group (n = 16) as more attractive) between the FULL VIDEO and the various components of the video

video and the various components for the NON-CONTRACEPTIVE group, in terms of proportions of raters who selected the high-fertility recording as more attractive, are reported in Table 2. Non-parametric correlations are used because of evidence that one of the sets of rater proportions was non-normally distributed (Shapiro-Wilk, $p < .05$); results are qualitatively identical with parametric Pearson correlations.

4.2.2 b Analysis by rater: ratings of NON-CONTRACEPTIVE group

The next topic of investigation was whether the raters in groups 1 and 3 (i.e. those who rated the MASKED FACE, FULL FACE, TORSO, SOUNDTRACK and SILENT VIDEO) selected the high-fertility recordings of the NON-CONTRACEPTIVE group more frequently in some modalities, using a mixed model ANOVA (rater as unit of analysis; within-subject factors: five stimulus types; between-subject factors: sex of rater; dependent variable: proportion of times that the rater selected the high-fertility recording as more attractive). Raters chose high-fertility recordings significantly more often than chance (test of the intercept: $F_{1,21}$ = 82.40, $p < .001$). Stimulus type was significant ($F_{4,84} = 5.29$, $p < .001$), and post-hoc comparisons with Bonferroni corrections for five comparisons revealed that raters selected the high-fertility FULL FACE more often than the high-fertility TORSO ($p = .040$), SOUNDTRACK ($p = .062$ – i.e. at borderline significance) and SILENT VIDEO ($p = .043$). Raters chose the high-fertility MASKED FACE more often than the high-fertility SOUNDTRACK ($p = .073$ – i.e. non-significantly) and did not differ in selection rates in relation to all other stimuli types (all $p > .16$). There was no effect of rater sex ($F_{1,21} = .006$, $p = .937$).

Analysis of the raters in groups 2 and 4 (i.e. those who rated the FULL VIDEO) with univariate ANOVA replicated these results: high-fertility videos of the NON-CONTRACEPTIVE group were selected significantly more often than chance ($F_{1,22} = 18.47$, $p < .001$). In contrast, in this group, there was a non-significant tendency for men to select larger proportions of high-fertility recordings than women ($F_{1,22} = 3.53$, $p = .073$).

There was evidence only for low levels of concordance between modalities within individual raters ($W = .173$, $p = .003$; unit of analysis: raters from groups 1 and 3 (i.e. those who rated the MASKED FACE, FULL FACE, TORSO, SOUNDTRACK and SILENT VIDEO) over five modalities). That is, there was a relationship of low effect size between the number of high-fertility recordings from the NON-CONTRACEPTIVE group selected by a rater in one modality compared with the number selected in another modality.

4.2.3 Discussion

The experiment set out to investigate the role of various modalities in the documented changes in ratings of female attractiveness over the menstrual cycle (review in Gangestad & Thornhill 2008). Non-users of hormonal contraceptives were recorded at high- and low-fertility phases of the menstrual cycle according to the backwards-counting method. The study set out to contrast ratings of these women with a control group of users of hormonal contraceptives who were similarly recorded on two occasions, scheduled according to their artificial 'cycle'. On each occasion, the women introduced themselves to a video camera, speaking for 20 seconds. One set of raters viewed the full video recordings ("FULL VIDEO") from the high- and low-fertility phases and selected the more attractive recording of each woman in a forced-choice paradigm. A separate set of raters chose the more attractive, for each woman, of the following recordings from the high and low fertility phase: a photograph of a neutral facial expression, cropped to conceal the hairstyle ("MASKED FACE"); a photograph of a natural facial expression ("FULL FACE"); a photograph of the clothing which would be revealed within the video (i.e. the body from neck to mid-thigh) ("TORSO"); the soundtrack from the video ("SOUNDTRACK"); and a muted version of the video ("SILENT VIDEO").

Overall, recordings from the high-fertility phase of the non-users of hormonal contraception were selected at greater proportions than recordings from the high-fertility phase of the users of hormonal contraception. Further, only recordings from the non-users of hormonal contraception were selected as the more attractive in the high-fertility phase significantly more often than chance. Previous research has shown that the high-fertility phase of the cycle is associated with increases in rated attractiveness in single isolated modalities such as the face, the voice, clothing, or in a static photograph (review in Gangestad & Thornhill 2008), and the present study extends this by finding these effects across different modalities.

There was no evidence that high-fertility selection rates varied more in some modalities than others when female participant was the unit of analysis. Visual inspection of the data suggests that raters may have been less able to pick the high-fertility recording when presented with the TORSO (which highlighted clothing choice) or SOUNDTRACK, and this was supported by a significant effect of stimulus type when the rater was the unit of analysis, i.e., for this particular stimulus set. Additional work is required to confirm whether these differences between modalities are robust across all women. Inconsistent with the results on TORSO ratings, previous work has linked the high-fertility phase to greater attractiveness ratings of clothing choice (Haselton et al. 2007; Schwarz & Hassebrauck 2008), and there are differences in the wearing of provocative clothing associated with differences in hormonal levels (Grammer et al. 2004). It is possible that changes in clothing choice would have been more easily detectable in the current study if women had been recorded in contexts where mate choice behaviour is more obvious; one study on clothing choice and the menstrual cycle (Grammer et al. 2004) recruited women at a night club, and another (Schwarz & Hassebrauck 2008), which had participants take photographs of themselves every day throughout the cycle, instructed women if possible to do this when dressed up to go out.

Likewise, in spite of the perhaps lesser effect of the SOUNDTRACK, existing research presents evidence for changes in vocal attractiveness across the cycle when women were recorded reciting a series of numbers (Pipitone & Gallup 2008). It may be that use of a video soundtrack, with variable semantic content, obscures changes in attractiveness that are specifically vocal. The present study did not ask raters to rate vocal attractiveness (with a standardised sentence) for three reasons. Firstly, pilot testing indicated that raters found this task much more difficult and frustrating than rating of the other stimuli types. Secondly, results from the literature employing objective testing of voice changes across the cycle has tended to present mixed evidence; voice change may be most apparent in professional singers or in subsets of women such as those experiencing symptoms of pre-menstrual syndrome (Silverman & Zimmer 1978; Abitbol et al. 1989; Higgins & Saxman 1989; Behr Davis & Lee Davis 1993; Abitbol et al. 1999; Chae et al. 2001). Thirdly, the SOUNDTRACK incorporates information on the voice, while also allowing for detection of changes of vocal style; for instance, there is also evidence for increases in verbal creativity (Krug et al. 1999) in the fertile phase.

The facial photographs replicated earlier work on attractiveness changes across the cycle (Roberts et al. 2004; Schwarz & Hassebrauck 2008), and presented the best evidence for detectable attractiveness changes. The use of masked neutral facial photographs and unmasked relaxed facial photographs meant that if there had been significant differences in selection rates between the two, it would not have been possible to distinguish whether these were due to effects of expression or effects of hairstyle. Indeed, the specific changes that provoke changes in facial judgments are unknown. The increase in basal metabolic rate around ovulation may provoke skin and lip coloration changes, as suggested by Roberts et al (2004). Likewise, the symmetry of soft tissue such as that found in the ears may reflect high-fertility (Manning et al. 1996). Women in the present study were instructed to remove facial make-up because make-up may obscure the link between attractiveness and cyclic hormonal levels (Law Smith et al. 2007). Future work might examine whether women are more likely to wear make-up at the high-fertility phase.

The hypothesis was that judges who viewed the extra information present in the FULL VIDEOS should be more likely to judge high-fertility recordings as more attractive compared with judges who viewed single components (Møller & Pomiankowski 1993; Rowe 1999; Candolin 2003). However, there was no evidence for this, suggesting indirectly that there were discrepancies between modalities in attractiveness changes. Indeed, when this possibility was tested directly, there was little evidence of concordancy across modalities. Further, high-fertility selection rates of the FULL VIDEOS correlated significantly and robustly, following correction for multiple comparisons, only with the SILENT VIDEOS, showing that a marked increase in overall personal attractiveness (in the full video) did not occur unequivocally in tandem with a marked increase in attractiveness in single modalities. The increased attractiveness associated with the high-fertility phase appears then to be apparent from different modalites in different women, rather than constituting a single "ornament".

It is as yet unclear whether attractiveness changes across the menstrual cycle constitute an evolved signal or are byproducts of hormonal changes (Gangestad & Thornhill 2008). If attractiveness changes constitute a signal, i.e. a structure that evolved in women with reference to an evolving response in men (Scott-Phillips 2008), then we might expect high-fertility to be equally apparent across modalities within one woman (Møller & Pomiankowski 1993; Rowe 1999; Candolin 2003), which we do not find in the present study. Alternatively, it is possible that there is some quality-dependent significance of the modality in which attractiveness changes are most apparent (see Johnstone 1996). It is possible that higher degrees of consistency across modalities would be observed if the rater is the female's partner, and hence more familiar with her appearance; or when the raters were attracted to the female targets, although we noted no robustly significant effect of rater sex on judgments; or in specifically sexual contexts, where explicitly prosocial behaviour may increase. In support of this last possibility, women at the high-fertility phase show greater pupil dilation in the presence of sexually significant stimuli (Laeng & Falkenberg 2007), suggesting a mechanism by which high-fertility could be recognised by a partner or potential partner. There could be other individual differences driving the magnitude of cyclic attractiveness changes, including relationship status or individual attractiveness; a larger study would be required to investigate these types of effects. Curiously, raters were significantly better at picking out the high-fertility recording of the non-users of hormonal contraceptives if these were made in the first session. We might speculate that women could be more likely to take steps to appear more attractive in conditions where they are most aware of their physical appearance, such as in a first and more stressful recording session, only when at high fertility. Together, the results suggest both that attractiveness changes are apparent in multiple modalities, and also that there is some conflict at the level of the individual woman between those different modalities.

4.3 General discussion

Design aspects of Experiment 4a and 4b, including the sample of rating targets, the sample of raters, the recording context, and the rating task, all deserve comment, and are dealt with in turn.

Both Experiments collapsed together a sample of rating targets who varied in aspects such as age, culture and race. In Experiment 4b, there were slightly larger proportions of women who did not describe themselves as 'white European' in the NON-CONTRACEPTIVE compared to the CONTRACEPTIVE group. This difference should not predict differences in appearance changes, although it is possible that it may have had a small conservative effect on the raters' ability to detect changes between the high and low fertility phases, as people are less sensitive to subtleties in appearance in other races (review in e.g. Meissner & Brigham 2001). The CONTRACEPTIVE group used a mixture of hormonal contraception types, and so would not be homogeneous in hormonal profile. Again, this does not affect the function of the control group, which is to allow for the comparison of ratings of women who are recorded twice in two separate sessions, controlling factors such as practice effects. Exclusion of the five women from the CONTRACEPTIVE group who did not experience menstruation gave rise to a similar pattern of results to that reported above, with a couple of minor variations: there was no longer a significant main effect of stimulus type in the initial analysis, and a couple of the reported non-significant tendencies disappeared. Such fluctuations are to be expected with reductions in sample size.

The groups of users- and non-users of hormonal contraceptives were very similar in terms of proportion of women in a relationship, although women were not asked whether they were sexually active. Sexual activity, and explicitly sexual contexts, might provoke bigger shifts across the cycle (c.f. Grammer et al. 2004). In Experiment 4a, age, sex and cultural factors may moderate the reported relationship between individual components and overall attractiveness. For instance, age affects rated attractiveness (Symons 1995), and so age may modify the relationship of a component to overall attractiveness, dependent upon the ease of age identification in that component. Rater age may also influence judgments; children, teenagers and young adults have been shown to give different weightings to facial and vocal attractiveness (Zuckerman & Hodgins 1993). Cultural differences that affect ratings of attractiveness and that are apparent across modalities such as apparel and accent may increase the correlation between individual components and overall attractiveness in culturally heterogeneous stimuli, and vice versa. The study did not attempt to separate out more biological factors such as body shape or voice quality from more culturally-influenced factors such as choice of clothing or speech, and although this meant that raters in both conditions were privy to identical information, future research might seek to examine the impact of these various possible influences. We note though the lack of impact of the additional semantic and personal information available in the soundtrack in Experiment 4a, as demonstrated by the significant, large-effect correlations between ratings of the speech and ratings of the standardised sentence. Finally, although our video stimuli were designed to emulate a first meeting, other modalities may also impact upon first impressions, including bodily odour (e.g. Li et al. 2007; Havlicek & Saxton in press) and bodily movement (Brown et al. 2005). Further research might look to investigate these avenues.

A fairly small number of raters was used in both Experiments, although this should not greatly affect results; attractiveness ratings between raters tend to be highly homogeneous (Rhodes 2006) and studies conventionally employ small numbers (e.g. Demattè et al. 2007). Raters encompassed both males and females, and those with both same- and opposite-sex sexual orientations. Differences in sexual orientation are unlikely to affect results of Experiment 4a, since all raters viewed targets as either objects of potential sexual attraction or potential sexual competitors. In Experiment 4b, only one participant (the homosexual male) may have been unlikely to judge the females with reference to standards of either potential mates or sexual competitors, but even if his ratings are systematically different, they would be unlikely to affect results due to the number of raters overall. Of potentially greater concern, the rating samples consisted of larger number of females than males. This sex weighting was mitigated somewhat in that the proportions of men and women who judged the components compared to the full videos was roughly equal. Experiment 4a found no effect of rater sex, but Experiment 4b found a non-significant trend for men to be better than women in selecting the high-fertility recording as more attractive. Research findings on the effects of rater sex on attractiveness judgments are highly variable. While in general terms men and women agree in their judgments of attractiveness (Rhodes 2006), there is also some evidence for systematic differences. Some studies have found that men are more consistent in their rating of similar cues across modalities (e.g. Lander 2008; Roberts et al. in review), although other work shows no effect of rater sex (e.g. Law Smith et al. 2007). Specific to studies on attractiveness changes across the cycle, one study (Roberts et al. 2004) found a female advantage in some contexts, whereas another (Haselton et al. 2007) found no effect of sex of judge. Men may be more practised at evaluating visual mate quality in women since information on female mate quality, such as health and fertility, is more likely to be visual than the corresponding information on male mate quality, such as resources (Symons 1995). A larger study could examine these questions, and also whether the fertility status of female raters bears upon their ratings of others (c.f. Fisher 2004).

The recording context, a self-introduction, was chosen to attempt to represent some aspects of a realistic zero-acquaintance mate choice context. Fully naturalistic recordings may strengthen some of the results. One way of increasing naturalism may be to employ a confederate to interact with the target participant, although this risks introducing interpersonal interactions between the confederate and the participant, such as the confederate's own judgments of the target.

The rating task took place in a group classroom setting. The presentation of rating targets on a screen rather than at individual computers should not have affected the ratings (Ekman et al. 1979). Raters in Experiment 4b viewed static images side by side, and may thus have been susceptible to left side bias; raters have been shown to select the left-hand face in a pair more often than the right (Klimkeit et al. 2003); and within a single face, the side to the viewer's left is more often used in judgments of face gender, emotion, or age and attractiveness (David 1993; Burt & Perrett 1997; Butler et al. 2005). Similarly, video and sound recordings may have been subject to recency bias: raters are more likely to evaluate an individual's overall performance positively if they have seen a good performance most recently (Steiner & Rain 1989), and incriminating evidence is more likely to lead to a guilty verdict by a jury if it is presented later in a trial (Costabile & Klein 2005). This is a problem inherent to all psychological testing of face or voice preference; in the present study, face pairings were reversed between the two sessions to attempt to counter this.

In Experiment 4b, the presentation of static photographs side by side, and presentation of vocal and video in sequence, follows standard presentation formats (e.g. Roberts et al. 2004; Jones et al. 2008b). Although presentation is then superficially different between modalities, at the cognitive level of information transfer it is arguably more equivalent than always using either side-by-side or sequential presentations. Switching between side-by-side and sequential presentation as appropriate allows for maximisation of information transfer; raters are not able to listen to two soundtracks, or take in dynamic information from two sources, simultaneously. Switching attention from one static image to the other, in contrast, does not detract from the amount of information that can be absorbed in each (i.e. the rater will not miss a sentence or a specific movement) and allows for accurate comparison.

Psychological research commonly asks raters to make judgments of attractiveness for a short-term relationship, a long-term relationship, or without specification of context; these affect ratings, as reviewed in Chapter 1. In the present studies, no context was specified. This allowed for comparison with existing literature on the relationship between single-trait and overall attractiveness (e.g. Peters et al. 2007; Lander 2008) and changes in attractiveness across the cycle (e.g. Roberts et al. 2004; Haselton et al. 2007). Further, the avoidance of the use of specific contexts allows raters to judge targets along dimensions most relevant to them; in particular, subjects occasionally are uncomfortable making judgments for short-term relationships, or feel that this is irrelevant to them. Future research might consider whether specifying contexts affects the relevance of the individual modalities. Such research should take into account the hypothetical background. For instance, if attractiveness changes across the menstrual cycle are supposed to function to allow the woman's partner to detect fertility, then judgments for a long-term context may be more relevant. This question of differences between short- and long-term judgments will be taken up in Chapter 5, yet this time women will be the subject rather than the object of the study.

Chapter 5: Preferences and trade-offs

5.0.1 The menstrual cycle and preferences for good genes

Women's mate preferences change systematically across the menstrual cycle, with the strongest preference for male markers of genetic quality at the conceptive phase of the cycle (review in e.g. Jones et al. 2008a). Researchers have suggested that these cyclic preference changes may sharpen female mate selection faculties, allowing women to select partners of the highest genetic quality when it is most biologically relevant to do so (e.g. Gangestad & Thornhill 1998; Penton-Voak et al. 1999a). Evidence that cyclic shifts in preference for markers of good genes are specific to mate selection, rather than being a more general effect of mood changes or person perception, comes from the findings that judgments of women's faces or voices are not affected cyclically in the same way as judgments of men's faces or voices (Johnston et al. 2001; Feinberg et al. 2006). Compellingly, these changes in preference for male attributes are responsive in a range of physical and psychological modalities, including the olfactory, visual, auditory and behavioural channels.

The first study to show an effect of menstrual cycle on female preference used male body odour, where the correlation between the rated attractiveness of the odour and the FA of the odour donor was examined at different cycle stages (Gangestad & Thornhill 1998). Although there was a high extent of agreement between the female participants over which body odours smelled more pleasant or more sexy, a significant and positive correlation between symmetry of the odour donor and ratings of odour pleasantness or sexiness only emerged from the ratings of the normally-cycling women at the fertile point of the menstrual cycle. This study was somewhat limited in that it examined the preferences of only a small number of women (*n* = 9 high-fertility, 19 low-fertility women), compared high- to low-fertility women rather than sampling the same women at different points in the cycle, and did not control for other factors such as the effects of MHC-similarity on odour preference. Nevertheless, the effect has been replicated with larger samples (Thornhill & Gangestad 1999b; Thornhill et al. 2003). The former of these two studies, and another (Rikowski & Grammer 1999), also found evidence for a link between male facial attractiveness and body odour attractiveness as judged by high- but not low-fertility women, but this effect would benefit from replication due to the lack of statistical significance or weakness of the effect size, the omission of statistical controls for multiple tests, and the absence of corresponding results in the study by Thornhill et al (2003).

Preferences for traits judged within the visual modality are also affected by the menstrual cycle. The original demonstration of this preference shift was in relation to decreased preference for male facial femininity at high fertility, particularly in the context of short-term relationships (Penton-Voak et al. 1999a), and this has since been replicated, and extended to apparent health in faces, and to male body sexual dimorphism, again particularly in the context of short-term relationships (Penton-Voak & Perrett 2000; Johnston et al. 2001; Jones et al. 2005a; Little et al. 2007b). Female preference for facial coloration may also vary cyclically (Frost 1994), although this study is hard to interpret due to its unusual methodology: females were compared in the first third versus the last two-thirds of the menstrual cycle rather than according to high versus low conception risk, and facial coloration was manipulated by developing a photograph for a longer amount of time, rather than restricting coloration difference to the facial skin; further, the difference between the evaluations of the groups of women only reached statistical significance in a one-tailed test. Not all 'good genes' indicators are preferred at high fertility; despite cyclic changes in female preference for the odour of symmetry reviewed above, and evidence for cyclic changes in perceptual response to visual symmetry (Oinonen & Mazmanian 2007), facial symmetry preference does not seem to increase at the conceptive phase (Koehler et al. 2002; Cardenas & Harris 2007; Oinonen & Mazmanian 2007).

In line with visual preference shifts, one study found that females in the most fertile phase of the menstrual cycle (but not those at other phases) showed a significant preference for lower- over higher-pitched male voices when judging the attractiveness of the voice in the context of a short-term relationship (Puts 2005). A second, within-subjects study (Feinberg et al. 2006) found that a preference for masculinised male voices (lowered fundamental frequency and decreased formant dispersion) was not restricted to females in the fertile phase, although the preference for masculinised voices did increase further for women during this time of peak fertility. The masculinised voices were also rated as more dominant, and indeed cycle-related preference for dominance has been demonstrated in the olfactory modality. Women at the most fertile phase, compared with females at other points in the cycle, rated the body odour of more dominant males (as measured by a standard dominance questionnaire) as more sexy (Havlicek et al. 2005).

In a study that set out to focus exclusively on behavioural preferences, Gangestad et al (2004) found that female judges at the most fertile point of the menstrual cycle, compared with females with a lower risk of conception, rated videotaped males as more attractive as potential short-term mates if their behaviour in that same video scored more highly on a set of behavioural measures seen by the authors to express strong 'social presence' (in an interaction with a female, these men tended to have a calm composure, to present themselves as athletic and avoid self-depreciation, and to maintain eye contact, amongst other things) and high levels of 'direct intrasexual competitiveness' (when competing with a male for a date with a female, these males might have specifically put down their competitor and engaged in minimal levels of laughter, for instance). It is possible that the participants were basing their preferences on physical as much as behavioural traits; the authors controlled for physical attractiveness as judged by a separate panel, but not physical attractiveness as rated by females of similar conceptive likelihood. Nevertheless, results of the study are in line with other findings of cycle-related preference changes. Finally, one other study to focus on behavioural aspects found that women expressed a preference for creative intelligence in men, such as that demonstrated by artists, at the high-fertility phase, when making judgments for short-term relationships (Haselton & Miller 2006).

5.0.2 The menstrual cycle and preferences for compatible genes

Cycle-related preference shifts have also been documented along a separate dimension, that of own-phenotype resemblance. The general effects of own-phenotype resemblance were reviewed in Chapter 1. To summarise, real-life couples demonstrate a degree of physical similarity (Zajonc et al. 1987; Bereczkei et al. 2002; Bereczkei et al. 2004), and experimental manipulation of visual similarity generally indicates that similarity enhances attractiveness (Penton-Voak et al. 1999b; DeBruine 2004; DeBruine et al. 2005), although one study found opposite-sex similarity to be aversive, at least in short-term relationship contexts (DeBruine 2005). Physical similarity may constitute an index of genetic compatibility or relative quality, possibly in part by providing information on MHC compatibility (Roberts et al. 2005b), and the choice of a physically similar partner may be adaptive (Bateson 1978, 1980, 1982; Read & Harvey 1991; Thiessen 1999). Menstrual cycle phase adds a layer of complexity in that women find own-phenotype resemblance to be more attractive at the low-fertility phase of the menstrual cycle (DeBruine et al. 2005; Jones et al. 2008a). Since this low-fertility phase shares hormonal similarities with pregnancy (reviewed in Jones et al. 2008a), researchers have suggested that own-phenotype preference may be an adaptation to seek out (supportive) kin during pregnancy (DeBruine et al. 2005; DeBruine et al. 2008; Jones et al. 2008a).

The existence of preferences for both absolute quality (relayed through cues such as high testosterone levels) and relative quality (referenced through, for example, own-phenotype resemblance), and the finding that the menstrual cycle affects both of these types of preference, suggests two related research questions. The first is how individuals weight absolute and relative quality in their mating preferences (Colegrave et al. 2002). In mice, females prefer to mate with males of good genes, and only base their choice of mate on MHC dissimilarity, denoting compatible genes, when there is very little variation in the genetic quality of the males, or when there is large variation between the males in the extent of their MHC dissimilarity (Roberts & Gosling 2003). The question of how humans weight absolute and relative quality has been little examined (Roberts & Little 2008). Probably the only relevant study in humans compared preference for facial averageness, indicating good genes, and facial own-phenotype resemblance, indicating compatible genes (Penton-Voak et al. 1999b). Participants were required to manipulate a face along a continuum from high to low self-resemblance. The mid-point of the continuum therefore represented a more average face. Participants' preferences did not differ significantly from the mid-point, showing that averageness is more salient in preference judgments than own-phenotype resemblance. As the authors note, an interesting extension would be to hold averageness constant, and to present participants with varying levels of own-phenotype resemblance for preference rating. Experiments 5a sets out to do this, using masculinity rather than averageness as the index of genetic quality. Women rated the attractiveness of male faces that had been manipulated along dimensions of sexual dimorphism and own-phenotype resemblance, in an experimental design recommended by Roberts & Little (2008).

The second research question, given that the menstrual cycle affects preference for both absolute and relative quality, is whether the cycle also influences how these are weighted. This is examined in Experiment 5b. Here, in a within-subjects design, women gave attractiveness ratings at the high- and low-fertility phase of the menstrual cycle to male faces that had been manipulated along dimensions of sexual dimorphism and own-phenotype resemblance, with ratings from users of hormonal contraceptives as a control.

5.1 Experiment 5a: Trade-offs in absolute and relative quality

5.1.1 Methods

5.1.1 a Female raters

Seventy-four Caucasian women were recruited from amongst university students and social contacts for a study on the stability of perceptions of attractiveness; participants were not told the specific study hypotheses. The majority of these women also took part in the studies reported in Chapter 4. At an initial meeting, women read a written description of the experiment and provided consent to participation. They were told that a photograph was required to be used to help in the understanding of their preferences, and were photographed directly facing the camera with a neutral expression. They then completed a questionnaire collecting basic demographic information.

5.1.1 b Stimuli creation

All stimuli images were of white individuals (aged between 18 and 25) with no spectacles or beards who had consented to the use of their facial image in experimental work. Photographs were taken under standardised lighting conditions and with participants posing with a neutral expression. To equate size, all images were aligned to standardise the position of the pupils in the image.

Composite images, composed of multiple images of different individuals, were used as base faces and as the basis for transforms. The composite faces were created using specially designed software (Tiddeman et al. 2001). Key locations (174 points) were manually marked around the main features (e.g. eyes, nose, and mouth) and the outline of each face (e.g. jawline, hairline). The average location of each point in the faces for each composite was then calculated. The features of the individual faces were then morphed to the relevant average shape before superimposing the averaged images to produce a photographic quality result. This technique has been used to create composite images in previous studies (see Benson & Perrett 1993; Tiddeman et al. 2001; Little & Hancock 2002).

A unique set of 54 male faces was created for each participant. These faces were derived from six composite male base faces, made using four individual images combined in the manner described above. In order to transform sexual dimorphism, two composites, one male and one female, were created. Each composite was derived from 50 individuals and each was made perfectly symmetric before transformation. Base faces were transformed on a sexual dimorphism dimension using the linear shape difference between the composite of 50 males and 50 females, following previous methods (see Benson & Perrett 1991; Perrett et al. 1998; Tiddeman et al. 2001). Image colours were not changed from the original. The transforms represented +50% masculinity and +50% femininity, based on the shape difference between these two composites. The transform gave rise to 18 images, composed of three images (one feminised, one unchanged, one masculinised) for each of the six base faces.

Following previous methodology (Penton-Voak et al. 1999b; DeBruine 2002, 2004), self-similarity was manipulated using the linear shape difference between feature points in the shape composite of 50 females against each participant's own particular shape. Image colours were not changed from the original. The transforms represented +25% self-similar and +25% self-dissimilar. This transform was applied uniquely to the 18 faces described above for each participant. Images were masked on the outline of the face so that hair and clothing cues were not visible. The final stimuli were then 54 masked faces for each female: six base faces by three levels of sexual dimorphism (feminised, original, masculinised) by three levels of self-resemblance (dissimilar, original, similar).

A transform of 50% sexual dimorphism was chosen because it meant that the images were still perceptually male when feminised, has been used in many previous studies of the effects of sexual dimorphism on face preference, and is known to affect judgments of attractiveness (Perrett et al. 1998; Penton-Voak et al. 1999a). A transform of 25% self-similarity was chosen in the aim of creating approximate perceptual equivalence with the 50% sexual dimorphism manipulation. There is more possible variability in the face shape of any one individual compared with the possible variability in the face shape of an average male or average female, meaning that a 50% transform towards or away from self-similarity could result in greater differences than a 50% transform along a sexual dimorphism continuum. Figure 6 demonstrates these manipulations.

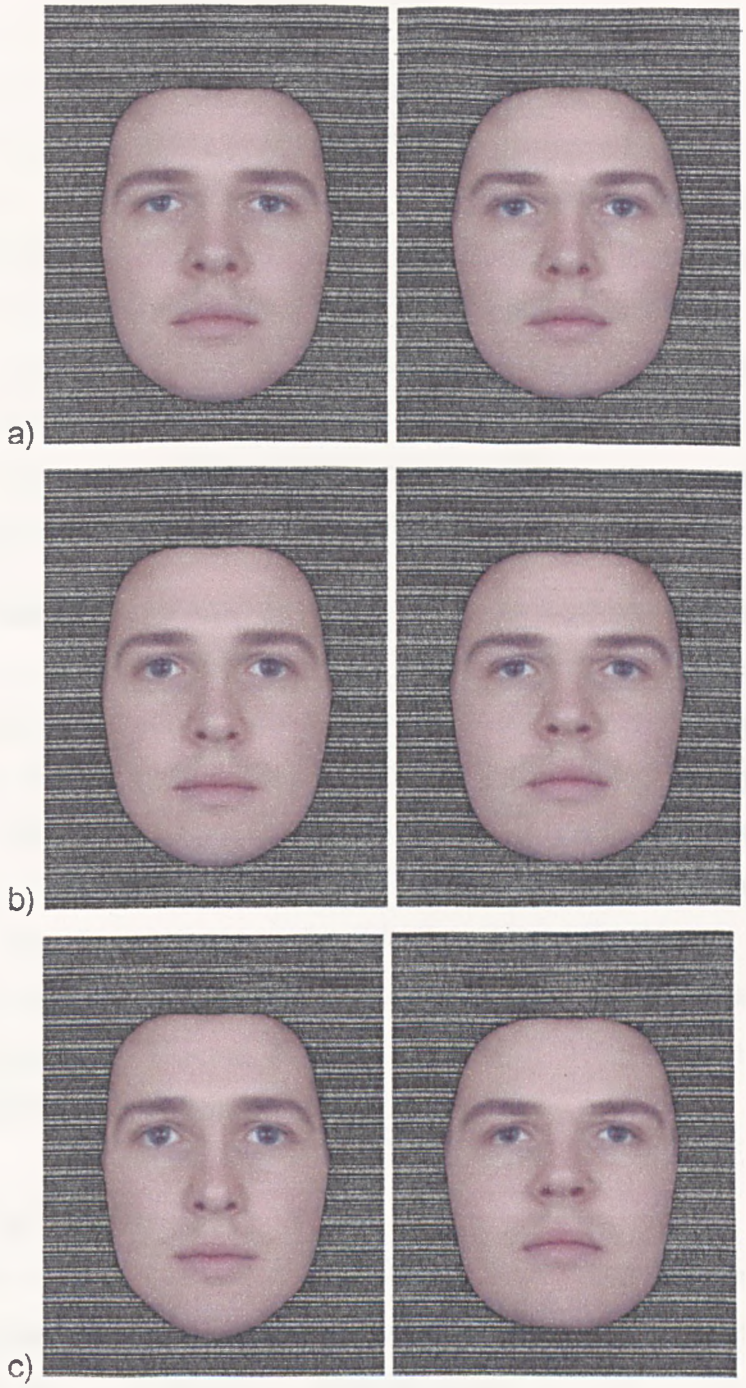

Figure 6. a) faces manipulated by (left) +50% masculine and (right) -50% masculine; b) faces manipulated by (left) +25% self-similar and (right) -25% self-similar; c) faces manipulated by (left) +50% self-similar and (right) -50% self-similar. Pairs a) and b) were used in the study.

5.1.1 c *Rating session*

Fifteen women did not attend the rating session, reducing participant numbers to 59, aged 16-39 (mean ± SD = 23 ± 5 yrs). At the rating session, women rated the set of face stimuli for attractiveness for a short-term relationship, and then rated the set of stimuli for attractiveness for a long-term relationship. Women were told that a short-term relationship might include a date or holiday romance, and a long-term relationship might include marriage or shared parenting. Stimuli were presented in randomised order using a Java applet.

Ratings were provided on a 7-point scale anchored by the verbal descriptors 'unattractive' and 'very attractive'. Following the collection of ratings, women were asked how much they knew about the study hypotheses. They were asked to talk generally, and also asked specifically what they thought the researchers were investigating. Around a third of the participants suggested that the faces were used to investigate responses to face manipulations, including size, shape and masculinity manipulations. No-one suggested that the faces may have been manipulated to resemble the rater.

5.1.1 d *Analysis*

Analysis was carried out in SPSS 15.0. Greenhouse-Geisser correction was used when there was evidence that data violated assumptions of sphericity (Mauchly's test, $p < .05$). If the analysis indicated a significant effect, pairwise comparisons (Least Significant Differences: i.e. uncorrected for multiple comparisons) were carried out. Mixed-model analysis with rater as unit of analysis was used in order to include the various levels (relationship term, self-similarity manipulation, masculinity manipulation) within the same analysis. To ease readability, the following terms are used to describe the manipulated facial images:

SD – the sexual dimorphism manipulation

SD-MASC – faces manipulated to increase masculinity

SD-ORIG – faces unmanipulated along the dimension of sexual dimorphism

SD-FEM – faces manipulated to increase femininity

SR – the own-phenotype resemblance manipulation

SR-HIGH – faces manipulated to increase self-resemblance

SR-ORIG – faces unmanipulated along the dimension of self-resemblance

SR- LOW – faces manipulated to decrease self-resemblance

5.1.2 Results

Repeated-measures ANOVA (three levels of SD, three levels of SR, short term or long term relationship) indicated significant main effects of both SD ($F_{2,116}$ = 16.10, $p < .001$) and SR ($F_{2,116}$ = 7.16, $p = .001$). Post-hoc pairwise comparisons showed that SD-FEM faces were rated significantly less attractive than SD-ORIG and SD-MASC faces (both $p < .001$) which themselves did not differ in their ratings ($p = .567$). Likewise, SR-LOW faces were rated significantly less attractive than both SR-ORIG ($p < .001$) and SR-HIGH ($p = .024$) faces, which themselves did not differ ($p = .213$). However, the main effects of SD and SR were modified by a significant interaction between SD, SR and relationship term

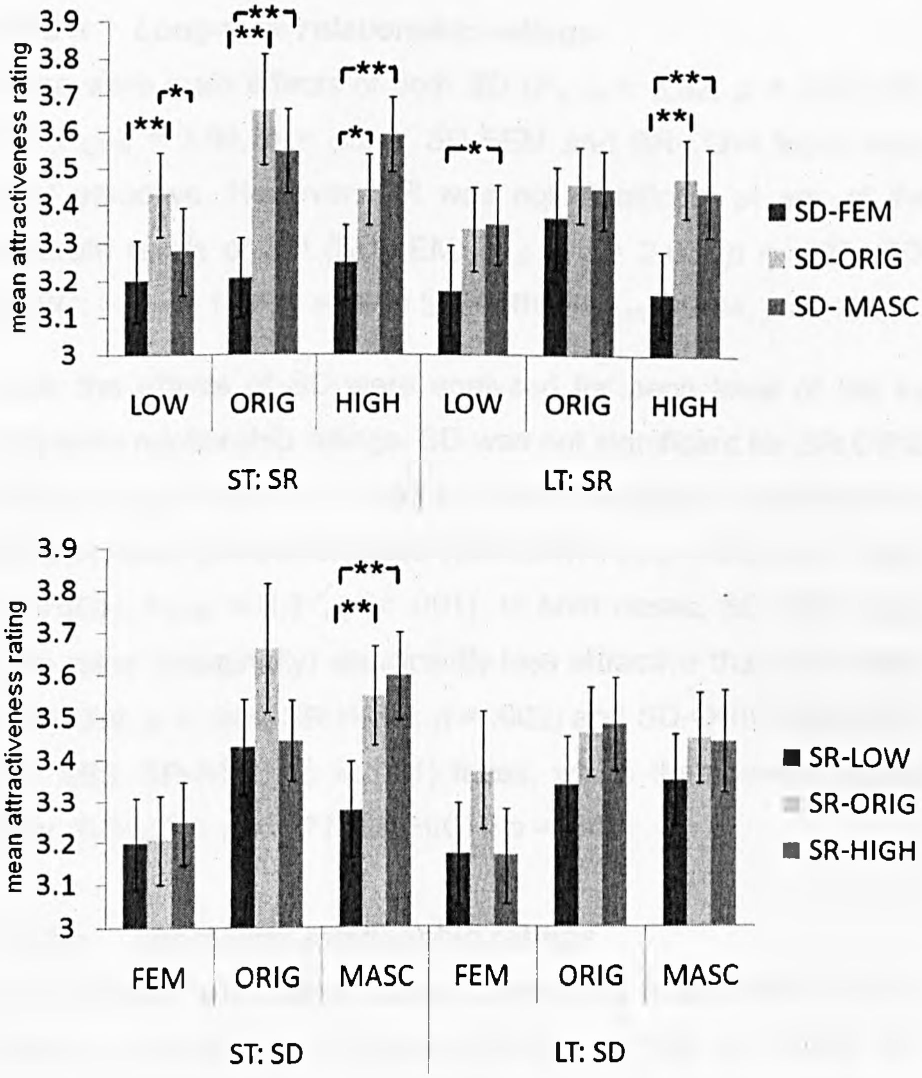

Figure 7. ST: short-term and LT: long-term relationship attractiveness ratings. Bars = mean rating ± SE; significant differences in LSD pairwise comparisons are shown at * p < .05 and ** p < .01

($F_{3.4,199.4}$ = 2.73, p = .037), which was analysed further, below. Results of these further analyses are represented visually in Figure 7.

5.1.2 a Long-term relationship ratings

There were main effects of both SD ($F_{2,116} = 6.32$, $p = .002$) and SR ($F_{2,116} = 3.65$, $p = .029$); SD-FEM and SR-LOW faces were least attractive. However, SR was not significant at any of the separate levels of SD (SD-FEM: $F_{1.5, 86.3} = 2.66$, $p = .090$; SD-MASC: $F_{2,116} = 1.00$, $p = .372$; SD-ORIG: $F_{2,116} = 2.04$, $p = .135$).

Next, the effects of SD were analysed for each level of SR for long-term relationship ratings. SD was not significant for SR-ORIG faces ($F_{1.6,93.5} = .46$, $p = .594$). SD was (marginally) significant for SR-LOW and SR- HIGH faces (SR-LOW: $F_{2,116} = 2.95$, $p = .056$; SR-HIGH: $F_{2,116} = 8.37$, $p < .001$). In both cases, SD-FEM faces were rated (marginally) significantly less attractive than SD-MASC (SR-LOW: $p = .048$; SR-HIGH: $p = .002$) and SD-ORIG (SR-LOW: $p = .055$; SR-HIGH: $p < .001$) faces, which themselves did not differ (SR-LOW: $p = .877$; SR-HIGH: $p = .642$).

5.1.2 b Short-term relationship ratings

In short-term relationship ratings, there were main effects of both SD ($F_{2,116} = 6.32$, $p = .002$) and SR ($F_{2,116} = 3.65$, $p = .029$); SD-FEM and SR-LOW faces were least attractive, although this was modified by a significant interaction between SD and SR ($F_{3.0,173.5} = 2.660$, $p = .050$).

Next, the effects of SR were analysed for each level of SD for short-term relationship ratings. In the analyses of both the SD-FEM and SD-ORIG faces, SR did not affect preference (in SD-FEM: $F_{2,116} = .19$, $p = .830$; in SD-ORIG: $F_{1.6, 93.8} = 2.07$, $p = .141$). In SD-MASC faces however, SR was significant ($F_{2,116} = 8.445$, $p < .001$): SR-LOW faces were liked significantly less than SR-ORIG (post-hoc pairwise comparison, p = .003) and SR-HIGH (p < .001) faces, which themselves did not differ (p = .582).

Separate analysis of the three levels of SR showed that SD was significant at all three levels (SR-LOW: $F_{2,116}$ = 4.49, p = .013; SR-ORIG: $F_{1.7,98.3}$ = 6.13, p = .005; SR-HIGH: $F_{2,116}$ = 8.57, p < .001). In SR-LOW, SD-ORIG faces were rated higher than both SD-MASC (post-hoc pairwise comparison, p = .043) and SD-FEM (p = .005) faces, which themselves did not differ (p = .346). In SR-ORIG faces, SD-FEM faces were rated lower than SD-MASC or SD-ORIG faces (post-hoc pairwise comparisons, both p = .002) which themselves did not differ (p = .472). Finally, in SR-HIGH faces, SD-MASC faces were rated near-significantly higher than SD-ORIG faces (post-hoc pairwise comparison, p = .052) which in turn were rated significantly higher than SD-FEM faces (p = .027).

The participants were recruited for Experiment 5b, and so some rating sessions were scheduled to coincide with the 'high-fertility' phase of the non-users of hormonal contraceptives. Accordingly, an unusually large proportion of women may have been at high fertility. To determine whether this was biasing the results, a second analysis was run where fertility phase (non-users of hormonal contraceptives at high fertility, n = 10) was entered as a between-subjects factor. There were no main effects of or interactions with this factor (all p > .1). Similarly, there were no main effects of or significant interactions with the between-subjects factor representing whether women were users or non-users of hormonal contraceptives (all p > .12).

5.1.3 Discussion

Fifty-nine women rated the attractiveness of male faces that had been manipulated simultaneously along dimensions of sexual dimorphism and own-phenotype resemblance in the contexts of a short-term and then a long-term relationship. The aim of the study was to investigate the different weightings of absolute and relative quality in women's mate choice decisions, and whether weightings differed by context (here, short- versus long-term relationship context), as we might predict (Colegrave et al. 2002).

Feminised faces were rated significantly less attractive than faces that had been masculinised or unmanipulated along the dimension of sexual dimorphism, in both short-term and long-term relationship contexts. This replicates previous findings that male facial femininity is often less attractive than male facial masculinity (DeBruine et al. 2006); further discussion of the attraction of facial sexual dimorphism was made in Chapter 1. Faces manipulated to look dissimilar from the rater were rated as less attractive than faces that had been manipulated to look more similar to the rater and than faces that were unmanipulated along the dimension of own-phenotype resemblance. This held for both short- and long-term relationship ratings. This replicates previous work on the attraction of own-phenotype resemblance, whether experimentally manipulated (Penton-Voak et al. 1999b; DeBruine 2004) or measured in real-life couples (Zajonc et al. 1987; Bereczkei et al. 2002; Bereczkei et al. 2004), but is inconsistent with work showing an aversive effect of experimentally manipulated own-phenotype resemblance in short-term relationship contexts (DeBruine 2005). The reason for this discrepancy is unclear, but seems unlikely to be a consequence of differences in the rating task. The current study used a Likert scale rather than forced-choice preference test, and a lesser degree of facial manipulation than (DeBruine 2005). However, other studies using forced-choice preference tests and greater degrees of manipulation have, like the present study, also shown positive effects of own-phenotype resemblance, albeit outside of the context of a short-term relationship (Penton-Voak et al. 1999b; DeBruine 2004). Instead, the discrepancy from previous findings could relate to local population differences. As reviewed in Chapter 1, individual differences give rise to differences in good genes preferences; such differences could also give rise to differences in compatible genes preferences in humans. For instance, in mice, preferences for compatible genes are affected by the range of genetic dissimilarity encountered in the population (Roberts & Gosling 2003).

Results also showed a significant three-way interaction between the two dimensions of manipulation (own-phenotype resemblance and sexual dimorphism) and the context in which ratings were made (short- or long-term relationship). In long-term relationship ratings, own-phenotype resemblance was more important than masculinity. Feminised faces were least attractive, but a robust effect of the sexual dimorphism manipulation (i.e. statistical significance at α = .05 both as a main effect and also in the subsequent pairwise comparison) was only demonstrated in the ratings of faces that had been manipulated to resemble the rater. As set out in Chapter 1, it has been argued that highly masculine men may not be a viable option for a long-term relationship for most women because they are too highly sought after (Little et al. 2001; Penton-Voak et al. 2003; Scott et al. 2008). Yet where both partners have a vested interest in developing a relationship (for example, by resemblance to each other), this may limit the marketplace and open up opportunity for women of lower genetic quality to partner more masculine men.

In contrast, in short-term relationship ratings, masculinity was more influential than own-phenotype resemblance. Own-phenotype resemblance only enhanced attractiveness ratings for masculinised faces, but masculinity raised attractiveness ratings at both medium and high levels of own-phenotype resemblance. Consistent with the results from a similar study in mice (Roberts & Gosling 2003), together these results suggests that raters applied a hierarchical or nested rule (Mays & Hill 2004), which varies according to relationship context. That is, in short-term relationship ratings, good genes (masculinity) are the more important, and compatible genes are only desired when good genes criteria are met. In contrast, in long-term relationships, own-phenotype resemblance is the more important; only at the highest level of self-resemblance does masculinity enhance ratings.

Faces that had been unmanipulated along dimensions of self-resemblance (SR-ORIG faces) tended to be rated as attractive as faces that resembled the rater, and faces that had been unmanipulated along dimensions of sexual dimorphism (SD-ORIG faces) tended to be rated as attractive as masculinised faces. This may be because unmanipulated faces may be perceived as more average and averageness is more important to ratings of attractiveness than own-phenotype resemblance (Penton-Voak et al. 1999b). The inclusion of unmanipulated faces in preference tests, along either the sexual dimorphism or the own-phenotype resemblance dimension, is unusual, and future work might consider redressing this.

The manipulations used were 50% sexual dimorphism and 25% self-resemblance. As discussed in 5.1.1 b above, these proportions were chosen to reflect previous work while maintaining approximate perceptual equivalence. Exact perceptual equivalence is unlikely, and indeed may vary by rater. Future research might seek to determine whether these results are equally applicable at other levels of facial transformation. Certainly, the inclusion of more than two levels of sexual dimorphism and own-phenotype resemblance seems important in light of findings of an asymptotic rather than linear function of own-phenotype resemblance on attractiveness ratings (Penton-Voak et al. 1999b). Previous research has used 100% (Penton-Voak et al. 1999b) and 50% (DeBruine 2004) own-phenotype resemblance manipulations; the present study demonstrates an effect at 25% manipulation.

5.2 Experiment 5b: Attraction and the menstrual cycle

5.2.1 Methods

5.2.1a Female raters

Seventy-four Caucasian women were recruited from amongst university students and social contacts for a within-subjects study on the stability of perceptions of attraction and attractiveness; they were not told the specific study hypotheses. The majority of these women also took part in the studies reported in Chapter 4 and 5a. Before women attended any study session, an initial meeting was arranged for each female volunteer, where she read a written description of the experiment and provided consent to participation. Females were told that a photograph was required to be used to help in the understanding of their preferences, and were photographed directly facing the camera with a neutral expression. Women then completed an initial demographic questionnaire which included questions on age, relationship status, ethnicity, hormonal contraceptive usage, date of last and next predicted menses, and details of average cycle length. Non-users of hormonal contraceptives all reported a usual cycle length of between 23 and 35 days. Users of hormonal contraceptives will be referred to as belonging to the CONTRACEPTIVE group, and non-users to the NON-CONTRACEPTIVE group.

5.2.1 b Session scheduling

Women attended two rating sessions. Sessions were scheduled as described in Experiment 4b. Women were contacted by email to seek information regarding onset of next menses subsequent to the sessions, and NON-CONTRACEPTIVE group women returned for a third (n = 3) or fourth session (n = 1) as necessary to obtain high- and low-fertility ratings from all participants. Since women could not always attend the laboratory at the appropriate time, five testing sessions were run over the internet; research has shown that internet-based ratings of opposite-sex attractiveness give equivalent results to lab-based rating sessions (Krantz et al. 1997; Epstein et al. 2001), and results of tests run over the internet are widely assumed to be equivalent to those run in the laboratory. Women were not excluded merely on the basis that they attended a recording session just prior to menstruation because research demonstrates large individual variability in the link between cycle phase and mood (Walker 1997); in any case, only one woman from the NON-CONTRACEPTIVE group final sample (see below) attended a recording session in the three days prior to menstruation. Four women used in the final sample (see below) did not experience regular menstruation. Of these four, one woman took oral contraceptives with a 28-day cycle and so the first day of the pill packet was defined as 'Day 1'. The remaining three women either used contraceptive implants (n = 1) or oral contraceptives with no 'Day 1' (n = 2). The sessions of one of these women were scheduled eight days apart, and her first session allocated to the 'low-fertility' phase and her second session to the 'high-fertility' phase. The sessions of the other two women were scheduled 14 days apart, and their first session allocated to the 'low-fertility' phase and their second session to the 'high-fertility' phase. Fifteen women did not attend any rating session, and 25 women did not carry out the task at both the 'low-fertility' and 'high-fertility' phase. The final sample consisted of 34 women (n = 13 NON-CONTRACEPTIVE). Nine women from the NON-CONTRACEPTIVE group and 15 women from the CONTRACEPTIVE group attended the first session in the 'high-fertility' phase. The rating task was described in Experiment 5a.

Following the final data collection session, women were interviewed by one of the female researchers (see Acknowledgments) about their understanding of the study hypotheses. They were given the opportunity to talk generally, and were also specifically asked why they thought they carried out the task on two separate sessions. Of the 34 women used in the final sample, two women from the NON-CONTRACEPTIVE group and 11 women from the CONTRACEPTIVE group suggested that the study may have been concerned with the effects of the menstrual cycle on preferences. Only two of the women from the CONTRACEPTIVE group remarked that their usage of hormonal contraceptives should eliminate this cycle effect on preference.

Analysis was carried out in SPSS 15.0. Greenhouse-Geisser correction was used when data violated assumptions of sphericity (Levene's test, $p < .05$). Repeated-measures analysis with rater as unit of analysis was used in order to include the various levels (relationship term, fertility status, self-similarity manipulation, masculinity manipulation) within the same analysis, while also allowing consideration of contraceptive usage.

5.2.2 Results

Mixed-model ANOVA (within-subjects factors: 3 x SD, 3 x SR, short term or long term relationship, high- or low-fertility; between-subjects factors: user or non-user of hormonal contraceptives; whether high- or low-fertility rating session was first) indicated a significant main effect of SD ($F_{2,62} = 8.288$, $p = .001$), but this was modified by a significant interaction between SD and relationship type ($F_{2,62} = 3.958$, $p = .024$). This significant interaction was analysed further with separate analysis of short-term and long-term relationship judgments.

SD was marginally non-significant in short-term relationship judgments ($F_{1.6, 50.8} = 2.719$, $p = .085$); feminised faces were given lower ratings than masculinised or unmanipulated faces, although this trend was non-significant in post-hoc pairwise comparison tests ($p > .07$). In long-term relationship judgments, sexual dimorphism level was significant ($F_{2,62} = 10.677$, $p < .001$); in post-hoc pairwise comparisons, SD-FEM faces were found significantly less attractive than SD-MASC ($p < .001$) and SD-ORIG ($p = .003$) faces, which themselves did not differ in the ratings they received ($p = .316$).

There were no other main effects or interactions (all $p > .09$), a pattern that was repeated when the sample was restricted to women in the NON-CONTRACEPTIVE group to maximise power to detect a fertility effect.

5.2.3 Discussion

Thirty-four women rated the attractiveness of male composite facial images that had been manipulated simultaneously across three levels of sexual dimorphism (masculinised, feminised, and unchanged) and self-resemblance (greater, lesser or unchanged resemblance to the rater) for short-term and long-term relationship contexts on a 7-point scale anchored by verbal descriptions ('unattractive' and 'very attractive'). Ratings were carried out at two points of the menstrual cycle, corresponding to high- and low-fertility phases for non-users of hormonal contraception, and the equivalent time periods with reference to the artificial menstrual cycle of users of hormonal contraception.

There were no significant effects of hormonal contraceptive usage or of fertility phase on ratings in Experiment 5a or 5b. This is inconsistent with most published previous findings (Penton-Voak et al. 1999a; Penton-Voak & Perrett 2000; Johnston et al. 2001; DeBruine et al. 2005; Jones et al. 2005a; Little et al. 2007b; DeBruine et al. 2008; Jones et al. 2008a). The within-subjects design of Experiment 5b amplifies the statistical power to detect an effect, and the sample sizes (Experiment 5a, $n = 31$ (10 at the high-fertility phase) and Experiment 5b, $n = 13$ non-contraceptive-users) compares favourably with at least one psychological investigation of changes across the menstrual cycle, which used 18 women, of which 11 were non-users of hormonal contraceptives (Miller et al. 2007). Nevertheless, the findings of no effect may represent a lack of statistical power due to small sample sizes. An adjusted design where women were paid to attend sessions, or where ratings were collected at regular weekly intervals over the course of a month, may have decreased drop-out due to non-attendance or to failure to obtain ratings during the fertile phase.

Medium-term variation: general discussion

Chapter 4 investigated the effects of the different modalities in judging appearance in general terms, and in judging appearance changes across the menstrual cycle. The first experiment showed that raters take account of both visual and aural attractiveness in their judgments of overall attractiveness. The second experiment found evidence that increased attractiveness at the high-fertility phase may be recognisable in the real world in holistic perceptions of women, and, surprisingly, that raters presented with information from a number of different channels simultaneously were not better at rating high-fertility recordings as more attractive, perhaps because there is discordance between modalities. Chapter 5 considered a corollary subject, namely, how women's preferences change across the menstrual cycle. Contrary to previous research, and possibly due to small sample sizes, it found no significant effect of menstrual cycle phase or hormonal contraceptive usage on preference. Nevertheless, in an experiment inspired by the previously noted cyclic changes in preference for cues of good genes and compatible genes, it showed that women's facial preferences can be described by a context-dependent hierarchical rule. This rule privileges compatible genes in judgments of attractiveness for long-term relationships, and good genes in judgments of attractiveness for short-term relationships.

The proximate mechanism driving changes across the cycle is unknown. Research on hormonal correlates of preferences have implicated a fairly large number of the hormones whose levels fluctuate across the cycle, including estradiol (Roney & Simmons 2008), oestrogen (Feinberg et al. 2006; Garver-Apgar et al. 2008), progesterone (DeBruine et al. 2005; Jones et al. 2005a; Puts

2006; Garver-Apgar et al. 2008), testosterone (Welling et al. 2007), and prolactin (Puts 2006). Changes in hormonal levels may ultimately both underpin, and also constitute biological constraints on (Via & Lande 1985; DeWitt et al. 1998), the adaptedness of cyclic changes.

While it is assumed that cyclic changes are biologically adaptive, the beneficiary of the changes is currently unknown. Many of the parameters that give rise to these changing preferences (physical masculinity, low vocal pitch, dominance) are positively correlated with testosterone levels which may constitute an honest signal of good genes, as set out in Chapter 1. Some researchers suggest that cyclic preference shifts may be useful to women to identify males of especially high genetic quality for extra-pair relationships (e.g. Little et al. 2007b). In support of this, preference shifts are sometimes only apparent when women make judgments for short-term but not long-term relationships, as reported in Chapter 5. Further, women of lower genetic quality themselves, who are thought to be more likely to adopt a short-term rather than long-term relationship strategy (Gangestad & Simpson 2000), show greater preference shifts: women with lower oestrogen levels showed greater fluctuations in their preference for masculinised voices across the menstrual cycle, while higher-oestrogen women did not vary their preferences so much as a function of fertility status (Feinberg et al. 2006). High-oestrogen women are more likely to have a more attractive body shape (Jasienska et al. 2004) and face (Law Smith et al. 2007), and thus oestrogen measurements may provide an assay of female quality. Yet this trend is not consistently found; for instance, menstrual cycle preference shifts appear less pronounced in more masculine, higher-testosterone women (Johnston et al. 2001; Scarbrough &

Johnston 2005). Further, preference shifts may function in established partnerships, as much as, or instead of, promoting extra-pair liaisons.

Indeed, the evidence from appearance changes across the cycle (Chapter 4) is more consistent with the idea that these function within established relationships rather than in extra-pair relationships. Unlike some other primates (Domb & Pagel 2001), women's changes in appearance are subtle. Attractiveness varies much more between women than within women, and in this way it seems highly probable that someone who was not intimately familiar with a woman would be unlikely to be able to detect attractiveness changes. It is true that the evidence that lap-dancers receive greater tips at the high-fertility phase (Miller et al. 2007) offers some contradiction to this; presumably, many of the tip-giving male customers would not have encountered the woman before. However, firstly, it is not clear whether tip increases respond to increased female attractiveness or, say, increased female diligence or ability to find high-paying customers. Secondly, female lap-dancers are likely to be much more homogeneous in attractiveness than the general population, and so cyclic changes in their attractiveness may be relatively more apparent. Future research on differences between partnered and unpartnered women, in particular in within-subjects designs, may help to answer some of these questions.

SECTION 3: LONG-TERM VARIATION: DEVELOPMENT DURING ADOLESCENCE

Section 3 examines the long-term changes in attractiveness judgments that take place during adolescence. Adolescence is a lengthy process, and as such constitutes an appropriate arena for the examination of long-term changes in attractiveness judgments. It also presents a theoretically exciting context for the investigation of attractiveness judgments because it is here when the human animal becomes capable of reproduction. Attractiveness judgments across adolescence have been very little examined; there are perhaps just two studies (Connolly et al. 2004; Saxton et al. 2006) that have addressed this topic. In this Section, attractiveness judgments are considered with reference to individual differences in biological change, encompassing both chronological age and pubertal development (Chapter 6), and also with reference to individual differences in experience (Chapter 7).

Many animal behaviours only emerge at the relevant point in the life stage of the animal, often alongside or as an outcome of the development of the relevant physical characters. The development of physical characters at the appropriate ontological stage is perhaps most noticeable in characteristics associated with courtship and intrasexual competition. Indeed, a test occasionally used to support the assertion that a trait is sexually selected is to examine whether it emerges at puberty (Andersson 1994; Cartwright 2000). This trend for capabilities to emerge as they are needed by the animal is, of course, not restricted to physical capabilities, but can extend to psychological capabilities. In humans, judgments of attractiveness are psychological capabilities that are thought to support biologically adaptive identification of high quality partners (Fink & Penton-Voak 2002; Rhodes 2006). We hypothesised that accuracy in these judgments may arise during adolescence, because the identification of mate quality is only biologically relevant when the individual becomes capable of reproducing.

Physical attractiveness judgments are perhaps most well understood in relation to facial attractiveness. In adults, averageness, symmetry and sexual dimorphism have systematic influences on judgments of facial attractiveness and are thought to be used to select desirable partners (reviews in Rhodes 2006; Roberts & Little 2008). Research on the attractiveness judgments of children has tended to focus on whether children respond in the same way as adults to faces that adults would describe as attractive, irrespective of how underlying traits such as symmetry may affect their judgments. In many ways, attractiveness judgments by children are not dissimilar from those made by adults. For instance, children who are given pairs of faces that have been pre-selected by adults to include an 'attractive' and an 'unattractive' face demonstrate preferences for the more attractive of the pair (Dion 1973), even at six months of age (Langlois et al. 1991). Another study found that adults and groups of children aged seven, 12 and 17 did not exhibit significant differences in their ratings of facial attractiveness (Cross & Cross 1971).

Yet these broad similarities between the preferences of different age groups may hide systematic changes in facial preferences during ontogeny. Concordance of agreement in the ordinal rankings of attractiveness of older children was found to increase in a between-subjects experiment comparing children aged between five and eight (Cavior & Lombardi 1973), and girls aged nine and 12 showed less pronounced preferences for attractive girls' faces than adults did (Kissler & Bäuml 2000). There are significant age-related differences in female assessments of the attractiveness of opposite-sex faces: pre-adolescent girls (aged 7 – 10), teenage girls (aged 12 – 15), and adult females (aged 20 – 34) all exhibited some degree of within-group concordance in their judgments of male facial attractiveness, but only judgments from the two older groups led to significant correlations between the facial and vocal attractiveness ratings of a group of men (Saxton et al. 2006).

The roots of facial preference may be apparent even in neonates. Infants preferentially track face-like objects 30 minutes after birth (Morton & Johnson 1991), and look longer at attractive faces from a few days after birth (Samuels et al. 1994; Slater et al. 1998). Yet facial averageness does not predict infants' judgments in the same way as it does adults': while infant looking time is thought to demonstrate preference, infants looked longer at less average and (non-significantly) at less symmetric faces, although this is possibly because of the documented infant preference for novelty (Rhodes et al. 2002). Accordingly, one of the topics tackled in Chapter 6 is an investigation of the presence or absence amongst adolescents of the preferences demonstrated in adults for facial symmetry, averageness and sexually dimorphic features.

Adults' judgments of vocal attractiveness have also been the subject of much recent research interest. Adult men prefer higher-pitched voices in women (Collins & Missing 2003; Jones et al.

2008b) and the faces of women with higher-pitched voices (Feinberg et al. 2005a), while adult women prefer lower-pitched voices in men (Collins 2000; Feinberg et al. 2005b; Feinberg et al. 2006; Saxton et al. 2006; Vukovic et al. in press). In contrast, a preference for higher-pitched voices is demonstrated by preschool children (Trainor & Zacharias 1998), and this preference is perhaps reflected in the tendency of adults to elevate their pitch in infant-directed speech (Fernald & Kuhl 1987; Kitamura et al. 2002). Likewise, one study found that pre-adolescent girls (aged 7 – 10) showed a non-significant tendency to prefer unmanipulated higher-pitched male voices, while the normal adult preference for lower-pitched male voices was only apparent from adolescent (aged 12 – 15) and adult raters (Saxton et al. 2006). It is clear that female attractiveness judgments of vocal pitch must undergo a fairly high-magnitude change during development, but the precise timing of this change is unknown, and again this question is addressed in Chapter 6.

Besides the theoretical basis for the hypothesis that preferences for cues of genetic quality should arise during puberty, there are also two distinct proximate mechanisms that may support these changes. The first derives from the tenet that familiarity is attractive (Zajonc, 1968); visual attractiveness judgments in particular are amenable to exposure effects (see e.g. Bereczkei et al. 2002; Little et al. 2003; Bereczkei et al. 2004; Little et al. 2005; Buckingham et al. 2006). Indeed, this mechanism has been previously demonstrated in the facial preferences of children. Due to their shorter stature, children will often see faces from below, thereby foreshortening the lower portions of the face being viewed. Consistent with their experience, pre-adolescents and younger children give higher attractiveness ratings to faces manipulated so that internal features are located lower than average, rather than higher than average, within the face (Cooper et al. 2006). As children go through puberty, their peers change in appearance, and one might predict changes in judgments of attractiveness that occur in response to the changing physical attributes of their peers. This would give rise to changes in judgments of physical attractiveness stratified by age or school class. To investigate this, the study included contrasts of the attractiveness preferences of groups of children of different ages (Chapter 6) and different levels of exposure to the opposite sex (Chapter 7).

The second proximate mechanism that could indirectly support changes in judgments of attractiveness during puberty is endocrinological. Physical attractiveness judgments change during the menstrual cycle in adult females, and these changes are associated with fluctuations in levels of various hormones (DeBruine et al. 2005; Jones et al. 2005a; Feinberg et al. 2006; Puts 2006; Welling et al. 2007; Garver-Apgar et al.; Roney & Simmons 2008). Although the physiological changes of puberty are of course distinct from those of the menstrual cycle, there is evidence for links between reproductive behaviour and pubertal hormones. Thus, the clinical condition known as precocious puberty, which stimulates the physical changes of puberty at a very young age, co-occurs with increases in sexualised behaviour (Thamdrup 1961; Ehrhardt & Meyer-Bahlburg 1994). Similarly, hormonal levels during adolescence correlate with sexual behaviour (Udry et al. 1985; Udry 1988; Halpern et al. 1993), and testosterone levels in adult men have been linked to differences in preference for sexual dimorphism in female faces (Welling et al. in press). Finally, early puberty has been linked to systematic differences in facial preferences in adults (Cornwell et al. 2006). If endocrine change is an important proximate mechanism of attractiveness judgment change during puberty, one would expect to see differences in judgments that correspond not to the age of the child but to his or her stage of pubertal development. This topic is also addressed in Chapter 6.

The studies reported in Chapters 6 and 7, then, set out to examine the trajectory and instigators of changes in adolescent judgments of facial and vocal attractiveness. Faces and voices were chosen as stimuli since there is a great deal of empirical and theoretical research on how facial and vocal characteristics affect their appeal in adulthood (reviewed in Chapter 1), and because they are both important in judgments of individual attractiveness (Chapter 4). Chapter 6 focuses on changes that occur in tandem with biological development, described by chronological age or by pubertal maturation. Chapter 7 in contrast considers how individual differences in experience during puberty, including experience of single- or mixed-sex schooling, and sex and number of siblings, may impinge upon preference.

Chapter 6: Biological development and attraction

Adolescence constitutes the ontogenetic stage between the dependency of childhood and the independence of adulthood. It constitutes both physiological changes, encompassing the changes of puberty when the child develops secondary sexual characters and becoming capable of reproduction (Timiras 1972), and socioculturally-constructed changes, when individuals take on the social obligations and responsibilities, and the legal rights, of adulthood. During this time, chronological years of experience, physiological (including neurological) development, and social expectations, lead to adult-like thinking, reasoning and behaviour replacing those of the child. Adolescence is associated with psychological changes including changes in cognition and perception, aggression and attention, emotion and motivation, and sexuality (Orr & Ingersoll 1995; Davison & Susman 2001; Dahl 2004). From a biological perspective, some of the most important changes relate to mate choice and reproduction, to which judgments of attractiveness form an essential prequel. The experiment described in this Chapter examines the relationship between biological development and judgments of attractiveness.

At the simplest level, biological development may be measured simply with reference to age. Age, rather than individual differences in pubertal development, is the best predictor of dating activity during adolescence, suggesting that age-related social norms and standards can explain at least some types of adolescent reproductive behaviours (Dornbusch et al. 1981). Yet differences in biological development during adolescence can also result from individual differences in pubertal maturation stage, something known to affect sexual behaviour (Udry et al. 1985; Udry 1988; Halpern et al. 1993). To link the changes of puberty to changes in attractiveness judgments, we must first understand broadly what puberty encompasses.

The development axes of puberty are threefold. Puberty encompasses the two neuroendocrinological changes of adrenarche and gonadarche (Grumbach 2002), and a third dimension of growth change. Adrenarche is the earliest stage of puberty, probably beginning from about age 6 or possibly earlier in girls, and a year later in boys (Parker et al. 1978; Cutler et al. 1990; Hubert & Carson 1990; Parker 1991). At this time, the adrenal glands begin to mature and to secrete adrenal androgens including dehydroepiandrosterone and its sulfate, androstendione, testosterone and estradiol (Sizonenko et al. 1976; Forest 1989; Parker 1991; Grumbach & Styne 1992). Early stages of adrenarche may be silent in terms of somatic markers of development (Dorn et al. 2006). Gonadarche, with reactivation of the hypothalamic-pituitary gonadal axis (Plant 1986, 2002), is accompanied by maturation of the ovaries or testes and the full development of secondary sexual characteristics (Dorn et al. 2006), precipitating spermarche in boys and menarche in girls. Gonadal steroids lead to female breast and pubic hair development, and to genital growth in both sexes (Dorn et al. 2006). Gonadarche begins from around age nine in girls and about a year later in boys (Grumbach & Styne 1992, 2003). The third set of pubertal maturational changes constitute changes in growth trajectory and in body size and composition (Reiter & Rosenfeld 2003), resulting from increased amounts of growth hormone (Dorn et al. 2006). It was originally thought that gonadarche was the more important instigator of sexual attraction (Money & Ehrhardt 1972), but adrenarche has recently received some attention in relation to its role in adulthood retrospective reports of initial feelings of sexual attraction (McClintock & Herdt 1996).

6.0.1 Introduction to the Experiments

Two experimental studies were carried out to investigate whether there was a consistent, directional relationship with age in relation to preference changes during adolescence. The first sampled face and voice preferences of British school children at two points, around a year apart. The second was a cross-sectional study of the preferences of Czech school children. Facial preferences were measured with forced choice preference tests of age-matched faces that had been manipulated along three facial dimensions thought to cue genetic quality (see Chapter 1) and as such are good contenders for cues that might be differently perceived before and after puberty: namely, symmetry, averageness and sexual dimorphism. Voice preferences were measured with forced choice preference tests of age-matched voices that had been manipulated for pitch, again a cue of genetic quality (Chapter 1).

Pubertal development data were also collected. The three developmental axes of puberty (adrenarche, gonadarche, growth change) allow pubertal development to be described with reference to a range of hormonal or somatic events. A wide variety of methods have been used in the literature, and indeed different research questions may merit different approaches to the quantification of pubertal stage (review in Dorn et al. 2006). The Tanner stages (Tanner 1962), assessed by an independent physician, are often regarded as the 'gold standard' in the measurement of puberty (Brooks-Gunn et al. 1987; Petersen et al.

1988; Carskadon & Acebo 1993; Taylor et al. 2001; Dorn et al. 2006), perhaps supplemented or, in particular clinical settings, replaced by measurement of boys' testicular volume (Hopwood et al. 1990; Dorn et al. 2006). Such methods are often inappropriate or unethical in non-critical or non-medical research, and questionnaire studies based on markers of somatic development may be preferred (Brooks-Gunn et al. 1987; Petersen et al. 1988; Carskadon & Acebo 1993).

Self-rated Tanner stages show good correlation with physician-rated Tanner stages (Duke et al. 1980; Morris & Udry 1980; Brooks-Gunn et al. 1987; Taylor et al. 2001), but this method still requires children to assess photographs of genitalia and secondary sexual characters, something that is not always acceptable to children or to their parents or guardians. A less intrusive method is the Pubertal Development Scale (Petersen et al. 1988). In this scale, children self-assess various physical traits from memory and indicate whether change from the child to the adult form has not yet begun, has just begun, is definitely underway, or is fully completed. Boys assess underarm and pubic hair growth, face hair growth, and voice change. Girls assess breast development, body hair growth, and also indicate whether and at what age menarche occurred. Both sexes assess skin change and growth spurt. The four stages are converted to a four-point scale, and the mean score used as a basis for comparison of pubertal developmental between individuals. Menarche is coded dichotomously (either 1 or 4 points). The Pubertal Development Scale has good internal consistency (Petersen et al. 1988) and has been independently validated to show high correlation with physician and self-evaluation of Tanner stages (Brooks-Gunn et al. 1987; Carskadon & Acebo 1993). Like Tanner stages, the Pubertal

Development Scale captures somatic markers of hormonal change. The growth of pubic hair is one of the earliest changes during puberty, and results from increased secretion of adrenal androgen (Hopwood et al. 1990). Increases in skin oils, leading to possible skin changes, is again a result of adrenal androgen secretion (New et al. 1981; Parker 1991). Peak height growth occurs in early puberty, around Tanner stage 2, in girls; and slightly later, in Tanner stage 3 or 4, in boys (Hopwood et al. 1990). Drawbacks of the Pubertal Development Scale include that it measures changes that tend to occur late in puberty, and conflates measures of adrenarche and gonadarche (Dorn et al. 2006). Gonadarche is more relevant for the age group in the study, but it is true that the Pubertal Development Scale, like all measures of puberty, is somewhat approximate of development. Further, questions about secondary sexual characteristics can still be considered offensive or inappropriate by the consent-granting body. A variety of methods of measurement of pubertal stage were used in the experimental studies described, dependent upon negotiations with the schools from which children were recruited.

6.1 Experiment 6: Attraction across adolescence

6.1.1 Methods

6.1.1 a *Stimuli creation*

All visual stimuli were created on the basis of facial photographs of 60 Caucasian children recruited in equal numbers from four groups (male or female; 11-13 (mean ± SD = 12:1 ± 0:6) or 13-15 (mean ± SD = 14:11 ± 0:11) years old) from local social groups or schools. Written parental consent and individual acquiescence was obtained from each participant. Conventional methods were used to create the facial stimuli (e.g. Perrett et al. 1994; Little et al. 2001). Facial features were marked out using dedicated software (Tiddeman et al. 2001) and used to create 12 sets (representing each level of: male or female; 11-13 or 13-15 years old; sexual dimorphism, symmetry, or averageness manipulations), each containing six pairs of images.

The mathematical averageness of stimuli was increased by adding 50% of the linear differences in 2D shape between individual images and the average shape for that category to six individual images from each sex and age category, allowing for a comparison pair with the original image. This method has been used previously to create stimuli for tests of responses to facial averageness (e.g. Jones et al. 2007). Following methods of previously published research (see e.g. Penton-Voak et al. 1999a; Little et al. 2001; Little et al. 2002; Jones et al. 2005a; Buckingham et al. 2006), stimuli which differed in sexual dimorphism were created by adding or subtracting 50% of the linear differences in 2D shape between the average face shape of the older boys and the average face shape of the older girls (i.e. 13 – 15 years old) to six photographs from each age and sex group. The face shape of the older children was used in creating both older and younger stimuli that differed in sexual dimorphism because pilot testing in adults revealed very little perceptual difference between images that had been masculinised and feminised using templates created from the younger stimuli (i.e. 11 – 13 years old). This is unsurprising given the low levels of sexual dimorphism evident in the faces of pre-pubertal individuals (Enlow 1990; Enlow & Hans 1996). Stimuli which differed in symmetry were created by first averaging six images from each of the four age and sex categories with their mirror image to produce a perfectly symmetric version, and then moving the image shape 100% towards (to increase symmetry) or 50% away from (to decrease symmetry) these perfectly symmetrical versions. Pilot testing in adults indicated little perceptual difference between the original and the 100% symmetrised images, potentially reflecting relatively low levels of asymmetry in children's faces (see e.g. Trivers et al. 1999). Symmetry manipulations created with reference to a continuum between symmetric and original faces are used in previous studies (e.g. Little et al. 2001; Little & Jones 2003). The same six faces were used for each manipulation type (averageness, symmetry, sexual dimorphism). Image colours were not adjusted from the original. Examples of the stimuli manipulations are given in Figure 8.

Recordings from six native English-speaking children from each sex and age group were used to make the vocal stimuli. The vowel sounds /oʊ/ (as in "go"), /u/ (as in "soon"), /a/ (as in "bar") and /i/ (as in "see") were recorded with a IM-DR420H 1-bit portable minidisc recorder (Sharp) at a sampling rate of 44.1 kHz and 16-bit quantization in a quiet room with a AT822 One point X/Y Stereo DAT microphone (Audio-Technica Limited, Leeds, UK) placed at a distance of around 20 cm from the speaker's mouth. Vowel sounds are conventional as stimuli in voice preference tasks (e.g. Collins & Missing 2003; Feinberg et al. 2005b) and allow easy perception of pitch while reducing speaker variation associated with intonation and articulation. Voice recordings were high- and low-pass filtered at 20 Hz and 7900 Hz to reduce non-vocal noise in the sound file while retaining the audible formant frequencies of children's vowel

Figure 8. Examples of image manipulation, applied to an adult base face (children's faces are not shown for reasons of consent). Top row: face has been masculinised (left) and feminised (right); middle row: face is original (left) and made more average (right); bottom row: face has been made more asymmetric (left) and more symmetric (right).

set	*stimulus*							
	1	2	3	4	5	6	mean	SD
younger girls	235	208	168	233	222	247	219	26
older girls	213	245	210	250	228	211	226	16
younger boys	226	218	222	240	136	193	206	34
older boys	133	107	172	132	151	128	137	20

Table 3. Mean pitch of voice recordings used as stimuli, followed by mean and standard deviation, for each stimulus set (Hz)

sounds (Jessica et al. 1999; Lee et al. 1999). All further acoustic measurements and manipulations were carried out using Praat 4.4.24 (www.praat.org). Fundamental frequencies were initially searched for between 180 and 700 Hz for the younger children (11-12 yrs old) and between 80 and 600 Hz for the older children (13-15 yrs old), and adjusted if necessary to take into account individual speaker variation. Original mean fundamental frequencies (before fundamental frequency manipulation) are set out at Table 3.

The PSOLA (Pitch-Synchronous Overlap and Add) method was used to shorten or lengthen vowels to obtain a duration of 0.35 s and then to create two new samples, one raised and one lowered by 20 Hz in fundamental frequency. Between-vowel silence was edited to occur at 0.5 s intervals, and amplitude normalised to 73 dB RMS. Finally, samples were converted from .wav to mp3 file format using the AllToMP3 Converter 1.6 (LitexMedia, Inc).

6.1.1 b *British raters*

Pupils were recruited from four schools charging similar levels of school fees, under approval from school senior management.

Parents were informed of the study by letter and provided with an opt-out form, which led to the non-participation of three pupils. The study took place over two school years. The attractiveness judgment tests were repeated with as many of the same pupils as were available plus a number of their classmates between nine and 13 months subsequent to the first test, subject to school time availability.

Children were given a slip of paper with a username that would be used to identify their answers in Year 1 and 2 of the study. Children filled in their name on the slip of paper, and the slips were stored securely by the school between Year 1 and 2 of the study. Children entered their username when supplying data. The schools were not given access to any data collected. In this way, the Year 1 and 2 data of the same child could be linked, but data remained anonymous, and neither the school nor the researchers could link data to any named child.

Stimuli were presented in pairs of faces or voices that were identical except for the manipulation applied (averageness, symmetry, sexual dimorphism, or pitch). Children had to indicate which of the two stimuli was more attractive and by how much. Children rated the stimuli created from children within their age group either at an individual computer (n = 229 Year 1; 213 Year 2) or provided pen-and-paper ratings of stimuli presented through an overhead projector and classroom stereo system (n = 102 Year 1; 122 Year 2). Such differences in presentation type are thought not to affect judgments; previous research has shown no differences in judgments of emotions, attitudes or personality traits dependent upon whether facial images were presented larger than life or about one-fifth life size (Ekman et al. 1979) and, more generally, facial attractiveness judgments are particularly robust to a range of presentation methods, remaining similar whether, for instance, faces are presented for 100 ms compared with unrestricted viewing times (Willis & Todorov 2006), or with whole versus half face information (Scheib et al. 1999). Further, face and voice rating data collected over internet websites are widely used in psychological research despite differences in visual image quality and size or aural amplitude etc arising from differences in individual computer screens or computer speakers, and face rating data collected in this way are considered to be equivalent to data collected under laboratory conditions (Krantz et al. 1997). Finally, presentation type was held constant within each set, such that each child viewed and made judgments of pairs of faces of equal sizes or voices presented in equal fashion.

To avoid overtaxing children's concentration spans, both same- and opposite-sex faces, but only opposite-sex voices (as voices took longer to evaluate), were rated. Stimuli presentation order and side were randomised. The younger children rated the same set of younger children's faces and voices in both Year 1 and Year 2; likewise, the older children rated the same set of older children's faces and voices in both Year 1 and Year 2.

Following the rating task, pupils provided basic demographic information including numbers of male and female children living in the same household and pubertal development. Year 1 questions were developed in consultation with the schools. Boys were asked whether they had underarm hair and whether their voice had broken, and girls were asked to indicate whether they had undergone menarche, and to rate their development in relation to their peers on a 4-point scale: 'much more developed', 'more developed', 'less developed', 'much less developed'. From these data, a three-point scale from the two self-reported measures of pubertal development (range: 0 – 2; boys: one point if his voice had broken and one point if he had underarm hair; girls: one point if menarche had been attained and one point if she rated herself more or much more developed than her peers). 150 boys and 111 girls answered all of these questions and were included in this analysis. In Year 2, pubertal questions were expanded to include the Pubertal Development Scale (Petersen et al. 1988) for all but one school where these questions were deemed inappropriate.

The data of any child who demonstrated extreme side bias in their preferences (Year 1 only, choosing consistently the image presented on one side 35 times out of 36; $n = 3$), or who entered an unrealistic year of birth (Year 1 only, $n = 2$), were excluded. Occasional technical problems or child lateness resulted in omission of some results, and children further had the option of skipping the rating of a face or voice. Mean preference score was only calculated if data from at least five judgments had been obtained following such omissions. Participant numbers and ethnicities following exclusions are as follows. Year 1: 89 boys and 74 girls in the younger class (mean ± SD = 11:10 ± 0:5 years; 122 Caucasian, 16 West Asian, 8 East Asian, 1 African, 16 non-respondents) and 93 boys and 75 girls in the older class (mean ± SD = 14:00 ± 0:6 years; 146 Caucasian, 6 West Asian, 5 East Asian, 1 African, 10 non-respondents). Year 2: 95 boys and 79 girls in the younger class (mean ± SD = 12:9 ± 0:4 years; 138 Caucasian, 15 West Asian, 8 East Asian, 1 African, 12 non-respondents) and 58 boys and 93 girls in the older class (mean ± SD = 15:00 ± 0:6 years; 131 Caucasian, 6 West Asian, 5 East Asian, 1 African, 8 non-respondents). Children who took part in both Year 1 and Year 2: 82 boys and 66 girls in the younger class (112 Caucasian, 15 West Asian, 8 East Asian, 1 African, 12 non-respondents) and 55 boys and 64 girls in the older class (101 Caucasian, 4 West Asian, 5 East Asian, 1 African, 8 non-respondents). Amongst the children who only took part in Year 2, 8 boys and 10 girls from the younger class and one boy and one girl from the older class omitted to give ages; they could not be included in any analysis where age is a covariate, and their ages do not contribute to the calculations of mean and standard deviation of age listed above.

6.1.1 c *Czech raters*

71 raters (*n* = 22 aged 12, *n* = 47 aged 13, *n* = 2 aged 14; 32 male) were recruited from three state schools in Prague in the Czech Republic and given Czech-language versions of the face and voice tests. Children only took part if their parent or guardian provided a signed a letter of consent to their participation. Face tests were presented with a Java applet which randomised face presentation order and side. Powerpoint was used to present the voice rating tests. Eight different versions (four female voice versions for male raters and four male voice versions for female raters) were created, each of which had identical numbers of lower-pitched voices presented first or second. Raters rated same- and opposite-sex faces from both the younger and older stimulus set, and opposite-sex voices from both the younger and older stimulus set. Children filled out a demographic questionnaire which included self-rated Tanner stages (Tanner 1962; Brooks-Gunn et al. 1987; Carskadon & Acebo 1993). Here, all children were given a series of five sex-specific images of pubic hair developmental stages to self-assess pubic hair development; this was converted

to a score on a five-point scale. Boys were given images of five stages of penis and scrotal development, while girls were given images of five stages of breast development; again, this was converted to a score on a five-point scale. For each child, the average of the two scores from the five-point scale was used as a measure of pubertal development (c.f. Vermeersch et al. 2008, who sum the two scores). Children were assured that they did not need to answer any questions, but that answers would be kept confidential and anonymous. They were seated in a quiet corner of the room away from the other children to complete the questionnaire. No child omitted any answer.

6.1.1 d *Statistical analysis*

Stimuli were presented in pairs of faces or voices that were identical except for the manipulation applied (averageness, symmetry, sexual dimorphism, or pitch). For each pair, the child had to state whether their preferred face or voice was 'much more attractive', 'more attractive', 'just more attractive', or 'guess', thus creating an eight-point scale from zero (if a child chose the more asymmetric, more distinctive or more masculine face or the higher pitched voice and found it 'much more attractive') through to seven (where a child chose the converse stimulus and found it 'much more attractive'). This method of calculating scores has been used previously (e.g. Buckingham et al. 2006). Three children who used the paper-based score sheets indicated which face/voice they preferred, but omitted to state the degree to which they preferred that face or voice (British children, Year 1: two children omitted degree for one face; one child omitted degree for one face and two voices). Their preferences were scored up as though they had chosen 'Guess', which allows accurate representation of their preference direction, while reflecting the preference degree that child usually chose (i.e. each of these children chose 'Guess' at least 30 out of a possible 36 times). Each child was presented with and made judgments on six stimulus pairs for each of the seven manipulations (male and female facial averageness, symmetry and sexual dimorphism; opposite-sex vocal pitch), and the mean score for each was calculated, giving rise to seven mean preference scores.

Statistical analysis was carried out in SPSS 15.0. Data were not always normally distributed, but the statistical methods applied are robust to this, particularly for large samples (Subrahmaniam et al. 1975; Olejnik & Algina 1984). The assumption of homogeneity of regression slopes was in some instances violated, but ANOVA is robust to such violations and loses little power when the groups are of equal sizes (Peckham 1968; Hamilton 1977). Greenhouse-Geisser correction is used when data violate the assumption of sphericity (Mauchly's test of sphericity, $p < .05$). Rater was unit of analysis since this allowed investigation of the effects of individual differences (e.g. age, sex, pubertal status) of the raters in terms of between-subjects effects, while also allowing consideration of how the raters responded differently to the various stimuli (i.e. comparing responses to symmetry with responses to averageness) while controlling for individual differences in rater ability.

6.1.1 e *Scale usage*

Scores may be confounded if one group of children uses more of the scale than another group. The child who uses the label 'much more attractive' four times to rate the feminine male face and twice to rate the masculine male face will have a score of ((4 x 7) + (2 x

0) = 28), while the child who makes the same decisions but uses the label 'just more attractive' will have a lower score of ((4 x 5) + (2 x 2) = 24, when in fact the children could vary only in their scale usage and not in their perceptions. If such differential scale usage were a confounding factor, we would find a significant effect of the standard deviation of a child's scores. There were no significant effects of any of the variables of interest (Year, sex, age) on the standard deviation of face or voice scores (face scores: mixed model analysis; within-subjects factor: Year; between-subjects factor: sex; covariate: age; all $p > .1$; voice scores: mixed model analysis; within-subjects factor: Year; covariate: age; all $p > .085$); further, inclusion of this standard deviation as a covariate in the analyses reported below has no substantial impact upon the results (i.e. results are qualitatively for the majority; in a couple of instances, results that are significant at conventional levels become significant at borderline levels i.e. $0.1 > p > .05$, or vice versa).

6.1.1 f *Side and order bias*

Because face presentation side was randomised, children differed with respect to the number of faces that they saw on the left- and right-hand side, or voices that they heard first or second. As discussed in Chapter 4, raters may demonstrate a left-hand visual selection bias (Nicholls et al. 1999; Klimkeit et al. 2003), and may show a recency bias in aural stimuli (Steiner & Rain 1989; Costabile & Klein 2005). This was of concern to the present sample because some groups of children were presented with the stimuli in classroom sessions, and since stimuli were randomised for presentation side there may have been an inadvertent bias in presenting one manipulation to the right or left (for faces), or first or second (for voices), possibly leading to systematic bias in the results relative to those groups of children. The analyses below have also been carried out with a between-subjects covariate to represent this. In analyses between Year 1 and Year 2, this covariate constitutes the number of times that the more average, symmetric or feminine face, or lower-pitched voices, was presented on the left in Year 1 subtracted from the number of times that it was presented on the left in Year 2, for the appropriate face category (i.e. for all faces overall if the analysis includes all faces; for male facial average manipulations alone if the analysis includes only male facial averageness manipulations; etc). In analyses of just Year 1 or just Year 2 data, this constitutes the number of times that the more average, symmetric or feminine face, or lower-pitched voice, was presented on the left. If voices are involved rather than faces, first and second are used rather than left and right. Since the inclusion of this covariate gave rise to qualitatively identical results, the results below are reported without this covariate for simplicity.

Items are referred to as follows.

adult-like judgment:

> higher ratings of femininity, averageness, and symmetry in faces, and of low pitch in male voices and high pitch in female voices

age (covariate):

> age of rater in months. Age at Year 1 is used for analysis of Year 1 data and all data together; age at Year 2 is used for analysis of Year 2 data

FACESEX (factor):

whether the stimulus rated is a male or female face

MANIP (factor):

whether the face stimulus was manipulated for averageness, symmetry or sexual dimorphism

PDS (covariate):

Pubertal Development Scale score

PUB (factor):

a three-point scale from the two self-reported measures of pubertal development (range: 0 – 2; boys: one point if his voice had broken and one point if he had underarm hair; girls: one point if menarche had been attained and one point if she rated herself more or much more developed than her peers). This is treated as a factor since it is an ordinal measure

SEX (factor):

whether the rater is male or female

stimulus type:

there are six facial stimulus types: the male facial averageness, female facial averageness, male facial symmetry, female facial symmetry, male facial femininity and female facial femininity manipulations. The two voice stimulus types are the pitch manipulations in male and female voices

TANNER:

the average of the two scores from the five-point scale (boys: pubic hair and penile/scrotal development; girls: pubic hair and breast development)

group – OLDER or YOUNGER (factor):

the younger children rated faces and voices of younger children, and the older children rated faces and voices of older children

Year (factor):

whether data are from Year 1 or Year 2 of the study

6.1.2 Results

6.1.2 a *Stimuli evaluation*

To test whether any stimulus was unusually affected by the manipulation applied, the average judgment scores from Year 1 were converted to z-scores separately for the male and female version of each face manipulation and voice pitch, and the resulting set of z-scores was checked for outliers. The z-scores fell within a normal distribution (no z-score with an absolute value > 1.96), indicating that the manipulations had similar perceptual effects on each stimulus.

Overall, in both Years, averageness, symmetry and femininity were preferred in male and female faces at levels greater than chance (single-sample t tests: all $t > 3.5$, $p < .001$). Children only rated opposite-sex voices; overall, boys preferred higher pitched girls' voices and girls preferred lower-pitched boys' voices (single-sample t tests: all $t > 3.8$, $p < .001$).

6.1.2 b *Do face judgments change with age? Longitudinal analysis, British raters*

Preferences for facial averageness, symmetry and femininity were apparent in both the younger and older groups of children, and in both Year 1 and Year 2 (single-sample t tests: all $t > 2.0$, $p < .05$). A mixed-design ANOVA (within-subjects factors: Year, MANIP, FACESEX; between-subjects factor: sex) was used to determine whether face judgments changed with age, between Year 1 and Year 2. There was no main effect of Year ($F_{1,241} = .56$, $p = .456$), but a significant interaction between Year and FACESEX ($F_{1,241} = 4.98$, $p = .027$). Separate analysis of male and female facial images (within-subjects factors: Year, MANIP; between-subjects factor: sex) showed no main effects of or interactions with Year in respect of male faces (all $p > .2$). Indeed, in Year 2, preference for male facial averageness and femininity had decreased non-significantly (male facial averageness, mean score 3.961 in Year 1 and 3.894 in Year 2; male facial femininity: mean score 3.942 in Year 1 and 3.878 in Year 2; male facial symmetry: mean score 3.703 in Year 1 and 3.747 in Year 2). There was a significant increase in adult-like judgment by Year for female faces ($F_{1,243} = 4.24$, $p = .041$, $r = .13$).

Next, the difference between Year 1 and Year 2 score was calculated for each child for each of the six stimuli types. Repeated-measures analysis of those scores with length of time between the two testing sessions as a covariate (within-subjects factors: MANIP, FACESEX; between-subjects factor: sex; covariate: length of time between the two testing sessions) showed a non-significant trend for a greater length of time between sessions to covary positively with greater increase in adult-like preference score ($F_{1,240} = 3.00$, $p = .085$, $r = .11$).

6.1.2 c *Do face judgments change with age? Cross-sectional analysis, British raters*

If the length of time between testing sessions is inadequate to see robust judgment changes in respect of male faces, then age-related judgment shift may only be apparent from these data in a between-subjects analysis. Accordingly, two separate mixed-design ANOVAs were carried out, one on Year 1 ratings and one on Year 2 ratings (within-subjects factors: MANIP, FACESEX; between-subjects factor: sex; covariate: age). In both analyses, greater age was associated with stronger adult-like judgments (Year 1: $F_{1,303} = 32.89$, $p < .001$; Year 2: $F_{1,301} = 23.78$, $p < .001$). In both analyses, this main effect of age was modified by a (near-) significant interaction between age and FACESEX (Year 1: $F_{1,303} = 6.04$, $p = .015$; Year 2: $F_{1,301} = 10.02$, $p = .002$) and between age, FACESEX and MANIP (Year 1: $F_{2,606} = 5.56$, $p = .004$; Year 2: $F_{2,606} = 2.78$, $p = .063$). Year 2 judgment scores were additionally modified by a significant interaction between age and MANIP ($F_{2,606} = 4.15$, $p = .016$).

To understand these interactions, 12 separate analyses (two Years x three MANIPS x two FACESEXES) on facial judgment scores collapsed across the two sexes (covariate: age) were carried out. There was an increase in score with age for averageness in female faces (Year 1: $F_{1,306} = 15.60$, $p < .001$; Year 2: $F_{1,304} = 4.61$, $p = .033$), averageness in male faces (Year 1: $F_{1,306} = 12.47$, $p < .001$; Year 2: $F_{1,304} = 34.46$, $p < .001$), femininity in female faces (Year 2 only: $F_{1,304} = 5.39$, $p = .021$), femininity in male faces (Year 1: $F_{1,304} = 33.61$, $p<.001$; Year 2: $F_{1,304} = 35.04$, $p < .001$), and symmetry in male faces (Year 1 only: $F_{1,305} = 5.33$, $p =.022$).

Next, the cross-sectional analyses were repeated, with the addition of group as a between-subjects factor (i.e. two analyses, one on Year 1 and one on Year 2 scores; in each, within-subjects factors: MANIP, FACESEX; between-subjects factors: sex, group; covariate: age). There were no main effects of, or interactions with, age (all $p > .3$). There was a non-significant trend for the older group of faces to reveal more adult-like preferences than the younger (Year 1: $F_{1,301} = 3.27$, $p = .072$; Year 2: $F_{1,300} = 3.08$, $p = .080$). There were also a number of significant interactions with group (Year 1: group * sex ($F_{1,301} = 4.14$, $p = .043$), MANIP * FACESEX * sex * group ($F_{2,602} = 5.60$, $p = .004$); Year 2: FACESEX * sex * group ($F_{1,300} = 8.57$, $p = .004$), group * MANIP * sex ($F_{2,600} = 2.84$, $p = .059$) that recommended further analysis. Post hoc tests contrasted the older and younger group separately for each face stimulus type, sex and Year, while controlling for age. Significant effects in Year 1 are that girls in the older group gave higher ratings of femininity to male faces ($F_{1,142} = 8.21$, $p = .005$). In Year 2, there was an effect of borderline significance for girls in the younger group to give higher ratings to symmetry in female faces ($F_{1,158} = 3.40$, $p = .067$).

6.1.2 d *Is there a confounding effect of face stimulus group? Czech raters*

Further examination of the question of whether stimuli in the two groups are equal as stimulus types was carried out by analysis of the ratings of the Czech children, who all rated stimuli from both younger and older groups. Mixed-design ANOVA (within-subjects factors: MANIP, FACESEX, group; between subjects factor: sex) showed that stimuli from the older group were given significantly higher ratings than the younger ($F_{1,67}$ = 8.12, p = .006), with no significant interactions.

6.1.2 e *Is there a confounding effect of practice?*

Analysis of Year 2 scores showed that the children who had carried out the task in the previous year (n = 267) gave significantly more adult-like judgments than the children who had not (n = 38) ($F_{1,300}$ = 4.26, p = .040; mixed-design ANOVA, within-subjects factors: MANIP, FACESEX; between-subjects factor: sex, participation in Year 1; covariate: age). There were no significant interactions with the factor participation in Year 1 (all p > .4).

6.1.2 f *Judgments of voice stimuli by age group*

Older girls demonstrated a significant preference for low pitch in boys' voices (single-sample t-test against chance levels, Year 1: t_{73} = 10.81, p <.001; Year 2: t_{90} = 11.01, p <.001). Younger girls showed a non-significant preference for boys' voices to be higher-pitched in Year 1 (t_{73} = -1.77, p = .081) but lower-pitched in Year 2 (t_{77} = .96, p = .341). Younger boys demonstrated a consistent significant preference for higher-pitched girls' voices (Year 1: t_{83} = 5.17, p <.001; Year 2: t_{93} = 4.69, p <.001), while older boys demonstrated a consistent non-significant preference for higher-pitched girls' voices (Year 1: t_{92} = 1.28, p = .206; Year 2: t_{48} = -.29, p =.774).

6.1.2 g *Do voice judgments change with age? Longitudinal analysis, British raters*

A repeated-measures ANOVA (within-subject factor: Year) was used to determine whether face judgments changed with age, between Year 1 and Year 2. Male and female voice judgments were analysed separately since raters only rated opposite-sex voices. Boys' preference for high pitch decreased over the two Years with marginal non-significance ($F_{1,123} = 3.47$, $p = .065$), while girls' preference for low pitch increased significantly over the two Years ($F_{1,125} = 4.64$, $p = .033$).

Next, the difference between Year 1 and Year 2 score was calculated for each child. There was no relationship between the length of time between the two testing sessions and the change in pitch preference (Kendall's tau = -.038, $p = .442$).

6.1.2 h *Do voice judgments change with age? Cross-sectional analysis, British raters*

Two separate mixed-design ANOVAs were carried out, one on Year 1 judgment scores and one on Year 2 judgment scores (covariate: age). These showed that greater age was associated with significantly decreased preference in boys for the higher-pitched female voices (Year 1: $F_{1,175} = 7.86$, $p = .006$; Year 2: $F_{1,135} = 4.90$, $p = .028$) and with a significant increase in preference by girls for the lower-pitched male voices (Year 1: $F_{1,146} = 58.67$, $p < .001$; Year 2: $F_{1,156} = 30.38$, $p < .001$).

This analysis of judgment scores from the two groups together assumes that stimuli in the two groups are equal as stimulus types because ratings of the younger stimuli by the younger children are combined in the same analysis as ratings of the older stimuli by the older children. To test this, the cross-sectional analyses were repeated, with the addition of group as a between-subjects factor (i.e. two analyses, one on Year 1 and one on Year 2 scores; in each, between-subjects factor: group; covariate: age). The inclusion of group removed all significant effects of age. Instead, there was a trend for older stimuli to be given higher ratings than the younger stimuli, although this was significant at conventional levels only in one analysis (Year 1 girls: $F_{1,145} = 5.76$, $p = .018$; Year 2 girls: $F_{1,155} = 2.91$, $p = .090$; Year 1 boys: $F_{1,174} = 2.61$, $p = .108$; Year 2 boys: $F_{1,134} = .67$, $p = .416$).

6.1.2 i *Is there a confounding effect of voice stimulus group? Czech raters*

Further examination of the question of whether stimuli in the two groups are equal as stimulus types was carried out by analysis of the ratings of the Czech children, who all rated stimuli from both younger and older groups. Repeated-measures ANOVA on girls' and boys' ratings separately (within-subjects factors: group) showed no significant differences between the ratings of the older and younger stimuli for either male or female voices (both $p > .11$; non-significant trend for low pitch to be rated higher in older girls' and younger boys' voices).

6.1.2 j *Is there a confounding effect of practice?*

Analysis of Year 2 voice scores showed no significant differences in voice scores between the children who had ($n = 128$ boys, 127 girls) and had not ($n = 9$ boys, 31 girls) participated in Year 1 (both $p > .6$; between-subjects factor: participation in Year 1; covariate: age).

6.1.2 k *Does pubertal development predict face judgments? British raters, Year 1*

Mixed model analysis (within subjects factors: MANIP, FACESEX; between-subjects factors: sex, PUB; covariate: age) showed a significant interaction between MANIP, FACESEX, PUB and sex

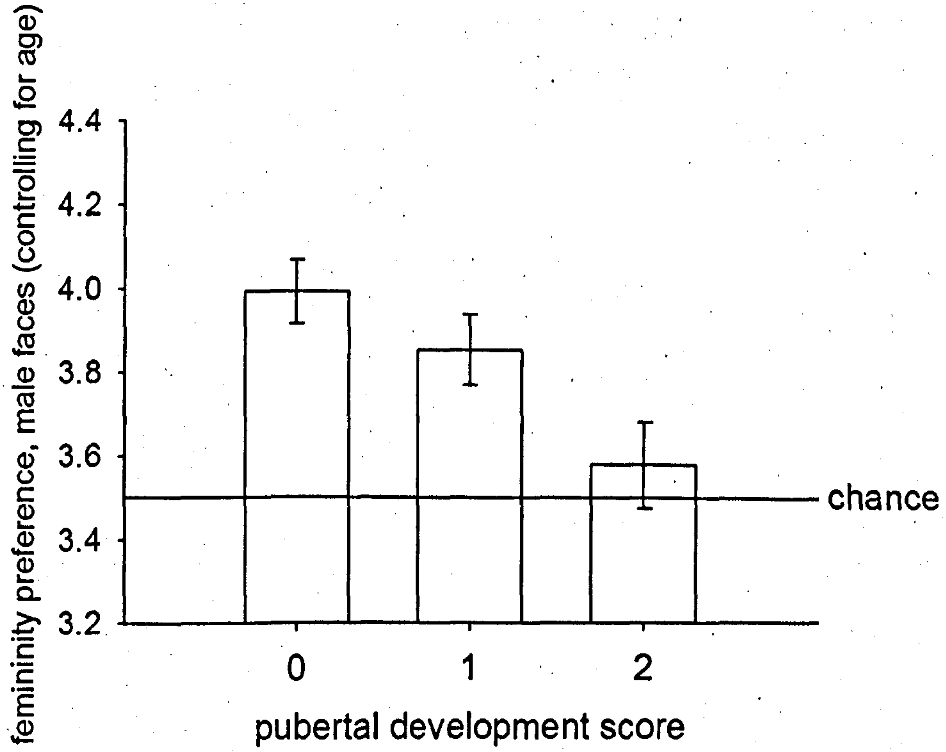

Figure 9. Boys' preference for femininity in male faces (estimated marginal means, controlling for age) by pubertal development score. Bars: mean +/- SE

($F_{4,506}$ = 2.60, p = .035). Separate follow-up analysis of each face stimulus type separately for girls and boys (between-subjects factors: PUB; covariate: age) showed only one significant effect: amongst boys, PUB increased as preference for femininity in male faces decreased ($F_{2,146}$ = 4.46, p = .013, Figure 9; all others p > .1). age remained significant in this analysis ($F_{1,146}$ = 8.10, p = .005), although if group is also included, only PUB remains significant ($F_{2,143}$ = 3.741, p = .026).

If boys and girls are analysed separately instead (mixed model analysis, within subjects factors: MANIP, FACESEX; between-subjects factors: PUB; covariates: age) then amongst boys, there is a significant effect of PUB ($F_{2,146} = 3.49$, $p = .033$): increased pubertal development is associated with decreased adult-like judgments. This result remains qualitatively identical with the inclusion of group in the model, and does not give rise to significant interactions with MANIP or FACESEX (both $p > .1$). There are no significant effects of or interactions with PUB in girls (all $p > .4$).

6.1.2 l *Does pubertal development predict face judgments? British raters, Year 2*

Mixed model analysis (within subjects factors: MANIP, FACESEX; between-subjects factors: sex, PUB; covariate: age) revealed no main effects of, or interactions with, PUB (all $p > .3$). The analysis was also carried out with PDS (as a covariate) substituted for PUB. There were no significant main effects of, or interactions with, PDS (all $p > .3$). Separation of boys and girls gave rise to qualitatively identical results, with the exception that in girls, increased PDS score was associated negatively but non-significantly with adult-like judgments ($F_{1,95} = 3.00$, $p = .087$), an effect that disappeared with the inclusion of group in the model ($F_{1,94} = 2.41$, $p = .124$).

6.1.2 m *Does pubertal development predict face judgments? Czech raters*

Mixed model analysis (within subjects factors: MANIP, FACESEX, group; between-subjects factors: sex, TANNER; covariate: age) revealed no main effects of, or interactions with, TANNER (all $p > .1$). Separation of boys and girls revealed one additional main effect: in boys, higher TANNER scores were associated with increased adult-like preferences ($F_{1,27} = 5.52$, $p = .026$).

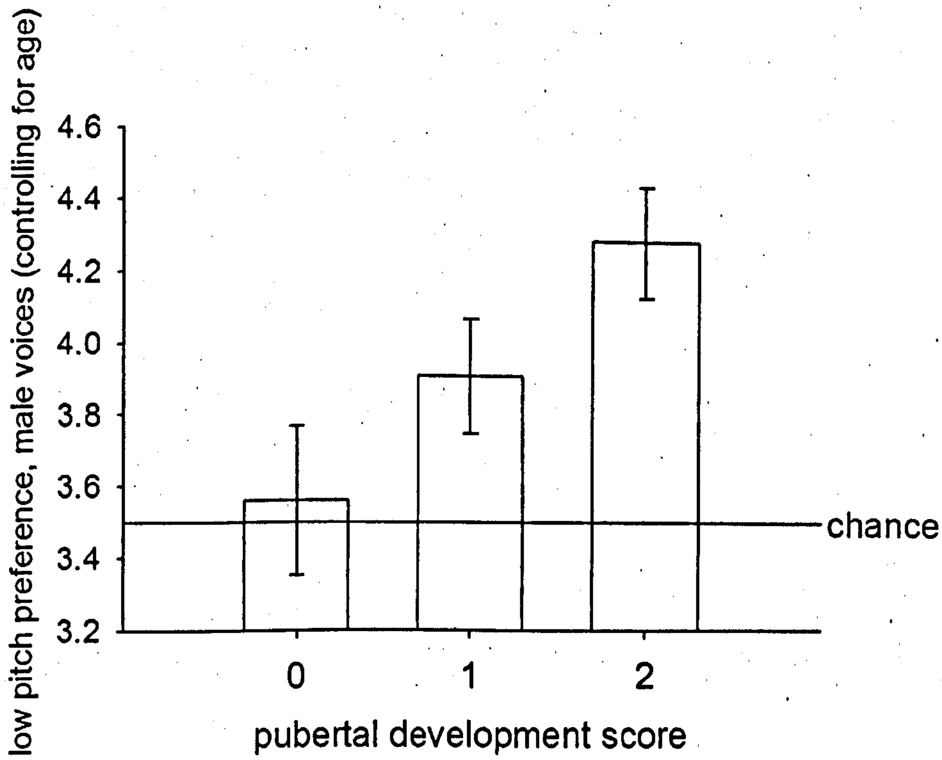

Figure 10. Girls' preference for low pitch in male voices (estimated marginal means, controlling for age) by pubertal development score. Bars: mean +/- SE

6.1.2 n ***Does pubertal development predict voice judgments? British raters, Year 1***

ANOVA of voice ratings for girls and boys separately (between-subjects factors: PUB; covariate: age) showed a significant increase in girls' preference for low pitch in male voices with increased PUB ($F_{2,107}$ = 3.33, p = .039, Figure 10); again, age remains significant in this analysis ($F_{1,107}$ = 14.795, p < .001), though the significant effects of age and PUB disappear with the inclusion of group in the model.

6.1.2 o *Does pubertal development predict voice judgments? British raters, Year 2*

ANOVA of voice ratings for girls and boys separately with either PUB as a between-subjects factor or PDS as a covariate, controlling for age, revealed no significant effects of PUB or PDS (all $p > .19$).

6.1.2 p *Does pubertal development predict voice judgments? Czech raters*

ANOVA of voice ratings for girls and boys separately (covariates: age, TANNER) showed no main effects of TANNER for girls ($p > .9$). Amongst boys, increased TANNER scores were associated with increased preference of borderline significance for high pitch in girls' voices ($F_{1,25} = 4.135$, $p = .053$).

6.1.2 q *Cross-modal preferences*

Girls' preference for male facial femininity was not correlated with preference for male low pitch voices either in Year 1 ($r = -.057$, $p = .500$, controlling for age) or in Year 2 ($r = .054$, $p = .502$, controlling for age).

6.1.2 r *Sex differences*

The initial analysis (within-subjects factors: Year, MANIP, FACESEX; between-subjects factor: sex) demonstrated that girls demonstrated significantly more adult-like preferences than boys overall ($F_{1,241} = 35.86$, $p < .001$), but this main effect was modified by a significant interaction between MANIP and sex ($F_{2,482} = 5.99$, $p = .003$). To understand this further, male and females were contrasted in separate analyses for each of the three MANIPS (within-subjects factor: Year; between-subjects factor: sex). Girls gave higher scores than boys to facial averageness ($F_{1,243} = 23.47$, $p < .001$), femininity ($F_{1,241} = 29.76$, $p < .001$) and, at borderline levels, symmetry ($F_{1,242} = 3.83$, $p = .051$).

This analysis assumes that the differences between boys and girls are equivalent in the younger and older groups. The analysis was rerun with the inclusion of group as a between-subjects factor (within-subjects factors: Year, MANIP, FACESEX; between-subjects factors: sex, group) to investigate this. Again, girls demonstrated significantly more adult-like preferences than boys overall ($F_{1,239} = 31.26$, $p < .001$), and this was again modified by a significant interaction between MANIP and sex ($F_{2,478} = 5.27$, $p = .005$), plus two new interactions with group: an interaction of borderline significance between MANIP, sex and group ($F_{2,478} = 2.69$, $p = .069$), and a significant interaction between FACESEX, sex and group ($F_{1,239} = 6.34$, $p = .012$). In light of these significant interactions, males and females were contrasted in 12 separate mixed-model analyses (six face stimuli types for the younger and then the older group; within-subjects factors: Year; between-subjects factor: sex). Girls gave higher scores than boys for averageness, femininity and symmetry in male and female faces. These difference between girls and boys was statistically significant in respect of femininity in older male faces ($F_{1,99} = 13.97$, $p < .001$), in older female faces ($F_{1,100} = 11.17$, $p = .001$), and in younger female faces ($F_{1,141} = 9.79$, $p = .002$); and in respect of averageness both in older male faces ($F_{1,100} = 8.87$, $p = .004$) and younger female faces ($F_{1,141} = 7.81$, $p = .006$). The difference between girls and boys was marginally non-significant in respect of symmetry in younger female faces ($F_{1,141} = 3.06$, $p = .083$) and averageness in younger male faces ($F_{1,141} = 4.08$, $p =$

.045), and non-significant in respect of the remaining contrasts (all $p > .14$).

6.1.3 Discussion

The study investigated how cues of mate quality were evaluated in faces and voices by peri-pubertal children, during the ontological stage when selection of mates according to their mate quality is becoming biologically adaptive. Raters aged 11 – 16 carried out forced-choice preference judgments of age-matched male and female faces that had been manipulated along the dimensions of symmetry, averageness and sexual dimorphism, and age-matched opposite-sex voices that had been manipulated for pitch, a sexually dimorphic vocal trait. Children were sampled from the year group that admitted children around age 11, and from a year group two years their senior, for children around 13 years. The rating task was carried out twice, at an interval of approximately one year. The study investigated effects of biological development, measured both by age and by pubertal development, on children's attractiveness judgments.

6.1.3 a *Age and face preference*

Both the younger and older groups of children rated the more average, more symmetric and more feminine male and female faces as the more attractive. To the author's knowledge, this is the first assessment of ratings of these facial parameters by this age group. Adults tend to give higher ratings of attractiveness to femininity in female faces, and facial averageness and symmetry in male and female faces, and all of these characters have been linked to genetic quality (review in Rhodes 2006). The finding that these characteristics are appealing to raters even before mate choice and mate competition are relevant (i.e. even in the younger group of children aged 11 – 13) adds to the evidence that these characteristics have non-sexual, social significance. Physical trait attractiveness has positive influences on social interactions in childhood (Dion 1973; Eagly et al. 1991; Langlois et al. 2000), and associations have been made between pro-social personality traits and facial symmetry (Shackelford & Larsen 1997; Fink et al. 2005) and facial femininity (Perrett et al. 1998). An awareness of how specific traits affect the attractiveness of faces may have social benefits even before mate choice judgments become relevant.

In this study, the children preferred feminine male faces. Adult preferences for both masculinised and feminised male faces have been reported (DeBruine et al. 2006; Rhodes 2006). Masculine male faces may cue indirect genetic quality (Folstad & Karter 1992), while feminine male faces may indicate positive personality traits such as warmth, honesty and cooperativeness (Perrett et al. 1998), leading to context-dependent preferences for male facial masculinity or femininity (Penton-Voak et al. 1999a; Little et al. 2001; Little et al. 2002; Penton-Voak et al. 2003). The finding that pre-reproductive children preferred male facial femininity may be in part due to the appeal and social relevance of male positive personality traits. In the older girls, a preference for male facial femininity could also reflect an age-related reproductive strategy. Raters who possess higher genetic quality show stronger preferences for facial masculinity (Little et al. 2001; Penton-Voak et al. 2003). Even the oldest adolescents in the sample must wait a number of years before they attain the age at which they are most desirable as a reproductive partner (Symons 1995), and as such their preference for femininity may reflect a strategic preference.

One could speculate that this preference for femininity in male faces among adolescents is also reflected in the popular sphere, in the characteristic boyish looks of male music bands appealing to younger teenagers.

The study presents some evidence for age-related preference change in respect of faces. Longitudinal analysis of Year 1 and Year 2 preference scores showed that children sampled for their preferences for averageness, sexual dimorphism and symmetry with the same set of faces on two occasions around one year apart gave significantly more adult-like preferences at the second sample in relation to female but not male faces.

It is possible that the lack of preference change in relation to male faces is a Type II error due to the short length of time between the two testing sessions (i.e. between nine and 13 months). This was tested in two ways. Firstly, the difference between Year 1 and Year 2 preference scores was calculated, and analysed with reference to the length of time between the two sessions. There was a non-significant trend for children whose testing sessions were further apart to show greater preference increases between the two sessions. This supports the notion that within this age group, adult-like preferences increase with age. However, since different schools were tested at different times, the possibility remains that this could be due to sessional effects: for example, academic or demographic differences between the pupils, or effects relating to the time of day of testing, or the position within the school term of the testing.

If the length of time between testing sessions is inadequate to see judgment development in respect of male faces, then age-related judgment shifts may only be apparent in a between-subjects, cross-sectional analysis. Accordingly, next, analysis of the effects of age on preference scores was carried out, separately for each data collection session. This allowed contrast of all of the children in the sample, ranging in age from 11 to 16. In this analysis, age effects were clearer. Children's preferences were significantly more adult-like with age in relation to averageness in both male and female faces, and femininity in male faces. There was also some indication of an age-related increase in preference for symmetry in male faces and femininity in female faces, but this was only apparent from the data collected in Year 1 and Year 2 respectively. It is not clear why judgments should increase in relation to some traits but not others in this way.

There is, however, reason for caution in the interpretation of these results, due to the use of the different stimulus sets for the different raters. To enhance ecological validity and appropriateness and to control for possible effects of own-age biases in perception (see e.g. Anastasi & Rhodes 2005), younger children rated the faces of younger children, and older children rated the faces of older children. The analysis of age effects assumes that the facial manipulations are perceptually equivalent in both younger and older stimulus sets. Yet the symmetry and sexual dimorphism manipulations act upon the degree of symmetry and sexual dimorphism of the base face. Sexual dimorphism increases during adolescence (Enlow 1990), and there is some evidence that bodily symmetry increases with age during adolescence (Trivers et al. 1999). This implies that there may be greater differences between the symmetric and asymmetric faces in the younger stimulus set, and that there should be greater differences between the masculine and feminine faces in the older stimulus set. The averageness manipulation acts upon both the degree of averageness of the base face, and also the qualities of the group from which the averageness template was obtained, and so differences in averageness levels between the older and younger groups (either age-related or random fluctuations) may give rise to greater or smaller perceptual differences between the older and younger groups.

Attempts were made to test this concern statistically by the inclusion of group in the between-subjects analyses (note that the effect of group is automatically controlled in the within-subjects analyses of preference change between the two Years, because children rated the same stimuli on both occasions). There was a non-significant trend for faces from the older group to provoke more adult-like judgments, and also evidence that group was affecting the different stimuli differently. Age effects were no longer significant in this analysis, but this is potentially because the effects of age and group cannot be adequately distinguished: both age and group may act similarly to increase adult-like preference, particularly as older children did not rate younger stimuli and vice versa, thereby disallowing separation of the effects of age and group (collinearity). Results from separate post-hoc analyses controlling for age were mixed: Year 1 girls in the older group gave significantly higher ratings to femininity in male faces compared with girls in the younger group, while in Year 2 girls in the younger group gave higher ratings to symmetry in female faces with borderline significance compared with girls in the older group. Finally, the Czech children who rated both younger and older stimuli demonstrated significantly more adult-like preferences in relation to the older stimuli.

In this context, it should also be noted that the face shape of the older children was used to manipulate both older and younger stimuli along the sexual dimorphism dimension because pilot testing in adults revealed very little perceptual difference between images which had been masculinised and feminised using templates created from the younger stimuli (i.e. 11 – 13 years old). This makes the assumption of something roughly equivalent to a linear trajectory of sexually dimorphic development between 11 – 13 and 13 – 15 such that the manipulation is not unnatural in the younger group; this has not been empirically tested, although work indicates that sexual dimorphism manipulations employing a range of techniques give rise to similar results (DeBruine et al. 2006).

It is possible that the finding of increased adult-like judgments in Year 2 is a mere practice effect, and this concern is supported by the finding that in Year 2, the children who took part in Year 1 gave significantly more adult-like judgments than the children who did not. However, this is not able to explain why the data show a more robust increase in adult-like preference in relation to female but not male faces yet no significant interaction between a child's participation and the sex of the rated face. It is possible that there was some systematic difference between the children who did and did not participate in Year 1, and it should be noted that inclusion of school type (single- or mixed-sex school; see Chapter 7) as a factor (data not shown) removed the significant effect of whether a child participated in Year 1, suggesting that uneven distribution of Year 1 non-participants between single and mixed-sex schools rather than practice effects may be responsible for these results.

Overall, the data present fair evidence that judgments become more adult-like between the ages of 11 and 16 in relation to female faces. The evidence that judgments become more adult-like between the ages of 11 and 16 in relation to male faces is not robust. There was fair evidence for an effect of stimulus type: older male and female faces both provoked more adult-like judgments than younger male and female faces. That is, faces with greater reproductive relevance are more likely to precipitate adult-like mating judgments. Female faces remain more neotenous than male faces during adolescence (Enlow 1990), and judgments of the attractiveness of male and female faces may thus follow different developmental paths. The response of the viewer to female faces may need to change more dramatically during ontogeny, because it can rely less on morphological changes in the female face to provoke appropriate mate value judgments. In this regard, it may be noted that at the neurological level, male and female faces appear to be categorised and processed differently (Little et al. 2005).

It seems unlikely that cognitive, motivational or concentration span differences could explain all of the differences between the age groups. Although older children may be cognitively better equipped for the task than younger children, children of all ages were able to carry out the task, as demonstrated by their directional preferences for at least some categories of stimuli. Maturation of self-regulatory skills and judgment continue to change well beyond the teenage years (Dorn et al. 2006), and extension of this study to older age groups may be warranted. Indeed, the present data were unable to replicate the existing finding that adults' preferences for masculine faces correlate with their preference for low-pitched voices (Feinberg et al. 2008), suggesting that a further period of maturation is required before the children in the sample have fully-developed mating judgments.

6.1.3 b *Age and voice preference*

Overall, girls preferred the lower-pitched male voices, and boys preferred the higher-pitched female voices. This reflects the preferences documented in adults: as reviewed in Chapter 1, low pitch tends to be preferred in male voices, and high pitch in female voices.

As the girls aged, they preferred lower-pitched male voices more. Indeed, a shift from a predominant preference for higher- to lower-pitched male voices in girls occurred in the data around the ages of 12 – 13. This finding is consistent with previous work that found a (non-significant) preference for high pitch in male voices amongst girls aged 7 – 10, but a (significant) preference for low pitch in male voices amongst girls aged 12 – 15 (Saxton et al. 2006). In the current study, evidence for this preference shift was apparent both from a longitudinal analysis, from comparison of pitch preferences in Year 1 and Year 2, and also from separate cross-sectional analyses of Year 1 and of Year 2 data. Inclusion of group (i.e. whether children rated older or younger voices) as a between-subjects factor in the cross-sectional analysis removed the significant effect of increased age, but again, this may be due to collinearity between age and the rating of the older stimulus set. The only significant difference in group was evident from girls' ratings in Year 1, where the older girls (who rated older voices) gave higher scores than the younger girls (who rated younger voices), controlling for age. Certainly, the Czech children who rated both stimulus sets did not give significantly different scores to the two sets. There was no evidence that increases in adult-like judgments were a 'practice' effect: the children who took part in both years did not produce significantly more adult-like judgments than the children who took part only in Year 2. Together, these results suggest that differences in pitch preferences with age result from changes in the perceptions of the rater more than being an effect of the stimulus set resulting from age-related changes in the attributes of the voices rated.

While acoustic differences in male and female voice pitch may begin from around 11 years of age (Lee et al. 1999) when male and female vocal tracts become anatomically distinct (Fitch & Jay 1999), boys' voices break at an age of around 13 to 14 years (Hollien & Malcik 1967; Tosi et al. 1976; Lee et al. 1999). Girls' preference for lower pitch therefore emerged around the time that some of their male peers' voices had dropped in pitch, indicating a role of exposure on preference judgments.

In the present study, the boys' preferences for higher pitch in female voices decreased with age, and while the younger group of boys preferred higher-pitched female voices, the older group did not show any significant directional preference. This contrasts with the established finding that higher-pitched adult female voices are rated more attractive (Collins & Missing 2003; Feinberg et al. 2005a). The pattern seen in boys' preferences may arise from their experiences with female voices. Girls' voices lower in pitch during childhood (Lee et al. 1999) and it is possible that the boys were using the lower pitches as a signal that the speaker was of older age and hence closer to sexual maturity. Further research could investigate whether additional experience is required before males show the adult-like preference for higher-pitched female voices, or whether such preference only emerges as other vocal cues such as size cues begin to indicate that the female is of sexual maturity.

6.1.3 c *Effects of puberty*

Prior to the survey of Year 1, a set of questions to measure pubertal development was developed in consultation with the schools recruited. Boys were asked whether they had underarm hair and whether their voice had broken, and girls were asked to indicate whether they had undergone menarche, and to rate their development in relation to their peers on a 4-point scale: 'much more developed', 'more developed', 'less developed', 'much less developed'. All of these measures bear upon pubertal development. Adolescent voice change and the growth of underarm hair in boys correlate with increasing levels of particular male hormones, including testosterone (Lee & Migeon 1975). Although menarche in girls not does necessarily correspond to reproductive potential, and often occurs without ovulation in adolescents (Apter 1980; Apter & Vihko 1983; Ibanez et al. 1999), it is a robust and key indicator of biological development, and corresponds significantly with height, trained raters' evaluations of physical development, and presence of other physical markers such as underarm hair and an adult female figure (Simmons et al. 1979). Overall physical development can also provide information on pubertal status, since the order and duration of some of the various physical changes (breast development, appearance of pubic hair, etc) differ between individuals (Marshall & Tanner 1969, 1970).

In the Year 2 survey, these questions were repeated, and the questionnaire was also expanded to include the Pubertal Development Scale. Following a change in management, one school restricted the questions on pubertal development allowed in Year 2. Year 1 and Year 2 data were analysed separately because the trajectory of pubertal development can vary (Hopwood et al. 1990) and so it would be inappropriate to assess the trajectory of face judgment change between Year 1 and Year 2 with reference to a static measure of pubertal development made at either Year 1 or Year 2.

Data on pubertal development were all self-reported. To maintain privacy and encourage truthfulness, children answered questions at an individual computer, or on an individual paper-based questionnaire. They were told that answers would be held confidentially and anonymously, and that they should leave blank any questions they did not want to answer. Despite these precautions, it is of course possible that some children were unable or unwilling to give accurate answers. Nevertheless, some significant effects of pubertal development were found.

Data on vocal preferences were analysed separately for boys and girls because children only rated opposite-sex voices. Data on facial preferences were analysed firstly with respect to all children together. This analysis was chosen to maximise the ability of the analysis to detect a sex-general effect of pubertal development irrespective of rater sex. Next, data on facial preferences were analysed separately in respect of boys and girls. This analysis was chosen to mirror the analysis of vocal preferences, concedes to the fact that the measures of pubertal development were different between the two sexes, and maximises the ability of the analysis to detect a sex-specific effect of pubertal development.

Amongst the British boys in Year 1, there was some evidence that greater pubertal development was associated somewhat with less adult-like facial judgments. This link was statistically significant and robust to the inclusion of group in the model. There was evidence both that this association between pubertal development and less developed adult-like judgments was apparent in relation to boys' judgments of all faces, and also that it was specific to boys' judgments of femininity in male faces. In Year 2, there was a non-significant, non-robust trend for increased development amongst girls as measured by the Pubertal Development Scale to be associated with less adult-like facial judgments. In contrast, amongst the Czech boys, increased pubertal development as measured by Tanner stages was associated with more adult-like facial and vocal judgments. Further, British girls' increased pubertal development was associated with more adult-like vocal judgments in Year 1, although this effect is not robust to the inclusion of group in the model.

These results are surprising since it was predicted that adult-like judgments should increase with pubertal development. Instead, the data reveal a mixed pattern, where pubertal development is sometimes associated with more adult-like judgments, sometimes with less adult-like judgments, and is sometimes without any association. Several explanations present themselves, although none is wholly satisfactory.

It is possible that pubertal development interacts with cultural or sex-specific expectations to lead to behavioural fluctuations. That is, amongst some groups of children such as the British boys, greater pubertal development could be associated with less attention to the task and consequent findings of lesser adult-like preferences. Early and late puberty in boys has been associated with increased non-normative behaviour during adolescence

(Williams & Dunlop 1999). However, levels of non-normative behaviour are dependent upon social and peer-group context (Rowe et al. 2004), which could in turn lead to the findings of different directions in the relationships between pubertal development and preference, and differences between the British and Czech children.

The various measures of puberty were capturing different aspects of development, and conflating measures of adrenarche, gonadarche, growth change, and the child's awareness of these changes. Each individual measurement could have an independent relationship with preference changes, and even vary in its relationship according to pubertal stage, thus leading to a relationship between development and preference in Year 1 but not Year 2, or conflicting relationships within Year. In Year 1, the positive correlation between boys' pubertal development (as measured by voice change and axillary hair) and their lower preference for femininity in boys' faces could result from their recent experience with their own developing faces, or indeed from enhanced attention to others at a similar developmental stage: serum testosterone levels increase with the onset of underarm hair growth (Lee & Migeon 1975), and there is evidence for a relationship between testosterone levels and male facial masculinity (Penton-Voak & Chen 2004; but see e.g. Swaddle & Reierson 2002). We did not collect data on menstrual cycle phase and so were not able to examine the known effects of cycle phase on preference (see Chapter 5) in this age group, although due to the high frequency of anovulatory cycles in the early postmenarcheal phase (Apter 1980; Apter & Vihko 1983; Ibanez et al. 1999) we would not necessarily expect to see a similar pattern as in adult women.

Preference differences may reflect an individual tactic or trade-off. Early pubertal development may reflect a reproductive strategy that is predicted to correspond to precocious sexuality, unstable pair bonds, and limited investment in child rearing (Belsky et al. 1991). Early puberty reported retrospectively has been associated with greater preferences in adulthood for opposite-sex facial sexual dimorphism (Cornwell et al. 2006). This association was not replicated in the present study, suggesting that the link between puberty and preference is moderated by a degree of complexity. The mixture in directions of puberty and preferences could arise if such tactics or trade-offs are undergoing development in the present sample.

In sum, although there was some evidence for a relationship between preference change and individual differences in pubertal development, the inconsistency of these findings suggest that the effect is not robust, or that the measures and tests used are not sufficiently sophisticated or well-targeted. This mixed picture is somewhat in line with existing literature: individual differences in pubertal development have been linked to individual differences in sexual activity (Udry et al. 1985; Udry & Billy 1987; Udry 1988, 1990; Halpern et al. 1993) but not dating behaviour (Dornbusch et al. 1981), and to facial preferences (Cornwell et al. 2006) but not waist-to-hip ratio preferences (Connolly et al. 2004). Ultimately, development of mating judgments and mating behaviour can be described by nothing less than an extremely complex interaction between individual differences in physical development, hormonal measures, and cultural and social controls (Udry & Billy 1987; Udry 1988; Halpern et al. 1993).

6.1.3 d *Sex differences*

Girls' face preference judgments were more adult-like than boys'. In general terms, girls tend to mature earlier than boys, to be more alert to socially relevant cues (Eagly & Wood 1985; Eagly 1987), and to have an advantage over boys in attention to facial cue judgments (McClure 2000). Accordingly, girls may be more practised in judging physical attractiveness. In addition, this could reflect greater choosiness by the girls. Good reproductive choice is more important for females than males because the cost of making a bad partner choice is much greater for females (Trivers 1972); females tend to be the choosier sex in judging partners (Andersson 1994) and this psychological sexual dimorphism may have driven some of the sexual differences noted in this peri-pubertal sample. It appears unlikely that differences in cognitive ability, motivation, and concentration span can explain all of the differences between the sex groups. It is possible that the difference between girls and boys is due to girls' greater motivation for the task, but if this is the case, this motivation to assess faces and voices may be the proximate mechanism that expresses the adaptive requirement for greater care by females than males in their choice of mates.

The study presents perhaps the first evidence that facial symmetry, averageness, femininity and vocal pitch have systematic influences on children's judgments, and that children's judgments of these parameters develop with age during the period at which mate choice judgments are becoming relevant. The results suggest that experience and exposure, and individual pubertal development, have measurable effects on the formation of face and voice pitch preferences during adolescence. Chapter 7 goes on to examine other external influences on these children's preferences.

Chapter 7: Effects of experience

Chapter 6 showed how individual differences in age and in pubertal development may give rise to systematic differences in the judgments of facial and vocal attractiveness. Yet individual differences at puberty are not restricted to those of biological development. Individuals also differ in terms of their experiences, and these may result in differences of judgment. The study described in this chapter investigated how differences in experience during adolescence of same-age faces of the same and opposite sex may impact upon preferences.

As reviewed in Chapter 6, the visual and auditory systems are responsive to experience, leading to familiarity preferences (Zajonc 1968). Face-based familiarity preferences have been previously demonstrated in children (Cooper et al. 2006, outlined in Chapter 6). Similarly, adults with greater levels of experience of children's faces demonstrate processing advantages in judgments of children's faces, and this advantage increases with the extent of their experience with children (de Heering & Rossion 2008). Chapter 6 also provided some evidence that exposure effects may not be merely visual, but may also extend to ratings of voice pitch. Exposure effects seem to be sex-specific: exposure to a certain type of male face will affect judgments of other male but not other female faces (Little et al. 2005). In light of all this, the present study investigated whether different levels of exposure to same-sex and opposite-sex peers, as experienced by children at single-sex compared to mixed-sex schools, or by children with different numbers of brothers and sisters, could impact upon preference judgments and the development of preference judgments. The hypothesis was that decreased levels of exposure to opposite-sex peers could hinder the development of adult-like judgments, and that this hindrance could be offset by the presence of opposite-sex siblings.

7.1 Experiment 7: Schools and siblings

7.1.1 Methods

Methods are as described in Chapter 6. Sample numbers are set out in Table 4.

As part of the demographic questionnaire, children were asked (Year 1) to state the number of boys or girls (aged 18 or under) who lived in the same household. Children were not asked to distinguish between full siblings and other children such as cousins, step siblings or adopted siblings who might live with them. This allowed children to keep private any personal family information; further, differences in relationship were not essential to the hypotheses. For simplicity, all reported co-habitees are referred to as 'brothers' and 'sisters' in the analysis below.

As explained in Chapter 6 ('Scale usage'), preference scores may be confounded if one group of children uses more of the scale than another group. Analysis of the standard deviation of voice judgments revealed a number of main effects of or interactions with school type. The results reported below are qualitatively identical with the inclusion of the standard deviation of voice judgments as a covariate, and so for simplicity, this covariate is not included in the analysis reported.

	younger children				older children			
	single-sex		mixed-sex		single-sex		mixed-sex	
	boys	girls	boys	girls	boys	girls	boys	girls
Year 1	45	43	39	31	45	38	32	35
Year 2	46	42	49	37	12	58	46	35
Both Years	43	40	34	26	12	33	28	29

Table 4. Sample sizes by rater school type, sex, and age group

7.1.2 Results

7.1.2 a School type and face preference

Preference scores were divided depending on whether they related to same- or opposite-sex judgments (i.e. collapsing together boys' ratings of boys' faces with girls' ratings of girls' faces, and boys' ratings of girls' faces with girls' ratings of boys' faces). School type did not have a main effect on face preference ($F_{1,238} = .81$, $p = .371$; mixed model ANOVA; within-subjects factors: year, MANIP, same- or opposite-sex judgments; between-subjects factors: sex, school type; covariate: age; results are qualitatively identical with the inclusion of group in the model). There were a number of significant interactions with school type. The highest-order significant interactions were between year, same- or opposite-sex judgments, school type and sex ($F_{1,238} = 9.03$, $p = .003$) which were followed up with separate school type contrasts for boys' and girls' facial stimuli judgments in Year 1 and Year 2, controlling for age and group (Table 5). Results are qualitatively identical with the exclusion of group, except that

		male raters		female raters	
		Year 1	Year 2	Year 1	Year 2
avg	m	$F_{1,156}$ = 1.42 p = .236	**$F_{1,140}$ = 4.30 p = .040: MS**	**$F_{1,142}$ = 4.41 p = .038: SS**	$F_{1,156}$ = 2.25 p = .135
	f	$F_{1,156}$ = 1.35 p = .247	$F_{1,140}$ = .17 p = .680	$F_{1,142}$ = .01 p = .944	$F_{1,156}$ = .39 p = .534
fem	m	$F_{1,156}$ = .55 p = .461	$F_{1,140}$ = 2.31 p = .131	**$F_{1,140}$ = 5.81 p = .017: SS**	$F_{1,156}$ = 2.15 p = .145
	f	$F_{1,156}$ = .10 p = .748	**$F_{1,140}$ = 7.97 p = .005: SS**	$F_{1,142}$ = 1.73 p = .191	**$F_{1,156}$ = 3.43 p = .066: SS**
sym	m	$F_{1,156}$ = 1.30 p = .256	$F_{1,139}$ = 2.08 p = .152	$F_{1,141}$ = .02 p = .881	$F_{1,156}$ = .15 p = .704
	f	$F_{1,156}$ = 1.12 p = .292	$F_{1,139}$ = 1.07 p = .302	**$F_{1,142}$ = 4.97 p = .027: MS**	$F_{1,156}$ = 1.33 p = .250

Table 5. School-type contrasts. Significant or borderline-significant results (p < .07) are indicated in bold, followed by the school type (SS: single sex; MS: mixed sex) where higher scores were given to stimuli manipulated for averageness (avg), femininity (fem) and symmetry (sym) in male (m) and female (f) faces

single-sex girls give higher judgments to averageness in male and female faces in Year 2 with borderline significance (p = .059 and p = .058 respectively).

There was also a significant interaction between MANIP and school type ($F_{2,476}$ = 8.43, p < .001). This was followed up with contrasts of school type for all three manipulations (mixed-design ANOVA; within-subjects factors: year, FACESEX; between-subjects factors: school type; covariate: age) which showed that children at mixed-sex schools gave higher ratings of facial symmetry with borderline significance ($F_{1,241}$ = 3.13, p = .078) whereas children at single sex schools gave significantly higher ratings of facial femininity ($F_{1,240} = 14.32$, $p < .001$). School type did not have a significant effect on ratings of facial averageness ($F_{1,242} = 1.25$, $p = .266$). Results are qualitatively identical with the inclusion of group as a factor.

The longitudinal data also allowed for investigation of any trajectory effects of school type. The difference between YEAR 1 and YEAR 2 score was calculated for each child and divided depending on whether they related to same- or opposite-sex judgments (i.e. collapsing together boys' ratings of boys' faces with girls' ratings of girls' faces, and boys' ratings of girls' faces with girls' ratings of boys' faces). Mixed-model analysis of those scores (within-subjects factors: MANIP, same- or opposite-sex judgments; between-subjects factors: sex, school type; covariate: age) showed a significant interaction between same- or opposite-sex judgments and sex ($F_{1,238} = 5.38$, $p = .021$) which was modified by a significant three-way interaction between same- or opposite-sex judgments, sex and school type ($F_{1,238} = 9.03$, $p = .003$). As there was no evidence for main effects of same- or opposite-sex judgments, but an interaction between same- or opposite-sex judgments and sex, this was followed up with separate analysis of the effects of FACESEX separately for boys and girls, contrasting school type. Girls' judgments of female faces became non-significantly more adult-like between the two testing sessions at single-sex compared with mixed-sex schools ($F_{1,125} = 3.16$, $p = .078$); at single-sex schools, girls' scores were non-significantly lower in Year 1 (mean score given to female faces, controlling for age; girls at single-sex schools: 3.884, girls at mixed-sex schools: 3.908) and non-significantly higher in Year 2 (mean score given to female faces, controlling for age; girls at single-sex schools: 3.995, girls at mixed-sex schools: 3.899). Boys' judgments of male faces became significantly more adult-like between the two testing sessions at mixed-sex compared with single -sex schools ($F_{1,114}$ = 5.89, p = .017); at mixed-sex schools, boys' scores were non-significantly higher in Year 1 (mean score given to male faces, controlling for age, boys at single-sex schools: 3.748, boys at mixed-sex schools: 3.810) and significantly higher in Year 2 (mean score given to male faces, controlling for age; boys at single-sex schools: 3.623, boys at mixed-sex schools: 3.813). There was no effect of school on judgments of female faces by boys or on judgments of male faces by girls (both $p > .17$).

7.1.2 b School type and voice preference

School type had no significant main effects on or interactions with voice preference (within-subjects factor: year; between-subjects factors: school type, group; covariate: age), but an interaction of borderline significance between school type and group in the girls' scores ($F_{1,121}$ = 2.91, p = .091). Separate analysis of the two groups showed that amongst the older girls, those at single-sex schools showed stronger preferences for low-pitched male voices than those at mixed-sex schools ($F_{1,58}$ = 4.73, p = .034). Analysis of Year 1 and Year 2 data separately replicated this finding in Year 1 only ($F_{1,71}$ = 7.00, p = .010).

The longitudinal data also allowed for investigation of any trajectory effects of school type. The difference between YEAR 1 and YEAR 2 vocal preference score was calculated for each child. ANOVA of those scores for boys and girls separately (between-subjects factors: school, group; covariate: age) showed no significant effects of or interactions with school type on preference change (all $p > .19$).

7.1.2 c Child co-habitees and face preference

A mixed-design ANOVA (within-subjects factors: year, MANIP, FACESEX; between-subjects factors: sex, school type; covariates: age, brothers, sisters) showed a significant interaction between FACESEX and brothers ($F_{1,227}$ = 5.24, p = .023): the presence of brothers tended to have a negative influence on ratings of female faces, and a positive influence on ratings of male faces. However, this was modified by a significant four-way interaction between year, MANIP, FACESEX and brothers ($F_{1,454}$ = 3.71, p = .025). Separate post-hoc analysis of each facial stimulus type for each year while controlling for age showed a negative association between brothers and Year 1 preferences for averageness in male faces ($F_{1,294}$ = 3.29, p = .071) and Year 2 preferences for femininity in female faces ($F_{1,255}$ = 6.38, p = .012). The factor 'group' was not included in the analyses reported since there should not have been systematic relationships between the presence of siblings and a rater's group; in support of this, results are qualitatively identical with the inclusion of 'group' in the models.

A mixed-design ANOVA (within-subjects factors: MANIP, FACESEX; between-subjects factors: sex, school type; covariates: age, brothers, sisters) on the set of scores representing the difference between Year 1 and Year 2 scores again showed a significant interaction between MANIP, FACESEX and brothers ($F_{2,454}$ = 3.70, p = .026). Post-hoc analyses revealed a positive association between brothers and the trajectory of judgments of male facial averageness ($F_{1,234}$ = 4.01, p = .046). Again, results are qualitatively identical with the inclusion of 'group' in the models.

7.1.2 d *Child co-habitees and voice preference*

A mixed-design ANOVA for boys and girls separately (within-subjects factor: year; between-subjects factor: school type; covariates: age, brothers, sisters) showed non-significant trends for brothers to covary positively with preference for higher pitch. This trend was of borderline significance if boys' and girls' preferences were analysed together ($F_{1,243} = 3.63$, $p = .058$). There were no effects of brothers or sisters on the difference between preference score in Year 1 and Year 2 (all $p > .3$).

7.1.3 Discussion

Children aged 11 – 16 carried out forced-choice preference judgments of age-matched male and female faces that had been manipulated along the dimensions of symmetry, averageness and sexual dimorphism, and age-matched opposite-sex voices that had been manipulated for pitch, a sexually dimorphic vocal trait. Children were sampled from the year group that admitted children around age 11 and from a year group two years their senior, for children around 13 years. The rating task was carried out twice, at an interval of approximately one year. Raters were recruited from a set of single-sex and mixed-sex private schools charging similar levels of annual school fees (i.e. between around £8 000 and £10 000) with the aim of reducing variance in responses attributable to socioeconomic background or other social factors (Cornwell et al. 2006; Swami & Tovée 2007) and thereby allowing comparison of mixed- and single-sex schools. The study investigated whether decreased levels of exposure to opposite-sex peers by attendance at single-sex schools could reduce the ability to detect cues of mate value in opposite-sex faces, and whether this hindrance could be offset by the presence of opposite-sex siblings.

7.1.3 a School type and face preference

There was no clear evidence to support the hypothesis that absence of opposite-sex peers at single-sex schools affected judgments of opposite-sex faces in a consistent fashion. Instead, school type had varying effects on judgments, dependent upon the sex of the rater and the type of face being rated. The clearest effect of school type on facial preference was in relation to facial femininity. Pupils at single-sex schools rated facial femininity significantly higher than pupils at mixed-sex schools. This was apparent from a variety of analyses, and was particularly evident in relation to girls' judgments, and in relation to female faces. Boys at mixed-sex schools and girls at single-sex schools demonstrated more adult-like judgments in respect of male facial averageness than boys at single-sex schools and girls at mixed-sex schools. Girls at mixed-sex schools demonstrated significantly stronger preferences for male facial symmetry than girls at single-sex schools, and children at mixed-sex schools gave higher ratings of facial symmetry with borderline significance. Girls' judgments of female faces became significantly more adult-like over the two year sampling period at single-sex compared with mixed-sex schools: at single-sex schools, girls' scores started non-significantly lower in Year 1, and ended non-significantly higher in Year 2. Boys' judgments of male faces became non-significantly more adult-like at mixed-sex compared with single-sex schools over the two year sampling period: boys' scores at mixed-sex schools started off non-significantly higher and ended significantly higher than boys' scores at single-sex schools.

Suggestions for reasons for these patterns can only be speculative. It may be tempting to suggest that girls at single-sex schools would be exposed to a larger proportion of female faces, and thereby would be more familiar with, and potentially give higher ratings to, feminine faces; yet other published work suggests that there should be no transfer between experience of female faces to ratings of male faces and vice versa (Little et al. 2005). Further, this advantage in rating same-sex faces is not found consistently in the other types of manipulation. There is perhaps a slight and inconsistent trend for girls at single-sex schools to have an advantage over girls at mixed-sex schools in judging cues of genetic quality, and for boys at mixed-sex schools to have an advantage over boys at single-sex schools. If this is a fair reflection of the situation, then it mirrors research suggesting that at both academic and social levels, girls may profit from single-sex schooling while boys may do better in mixed-sex classes (Jimenez & Lockheed 1989; Burgess 1990; Haag 1998; Jackson 2002), and these differences could thus reflect more general differences in concentration or attention to the rating task.

7.1.3 b Child co-habitees and face preference

The presence of brothers was associated negatively with Year 1 preference for averageness in male faces, and positively with the difference between Year 1 and Year 2 scores of averageness in male faces. This effect is difficult to explain. The presence of brothers was associated negatively with Year 2 preference for femininity in female faces. Again, although it is tempting to suggest that raters with brothers may have been more familiar with and hence rated more highly the more masculine girls' faces, previous work suggests that judgments of male and female faces do not communicate in this way (Little et al. 2005). There were no effects of sisters on facial judgments. Girls' faces remain more neotenous during development compared with boys' faces (Enlow 1990;

Enlow & Hans 1996), and so the changes that raters would see in same-age female faces may be less salient.

7.1.3 c School type, child co-habitees and voice preference

Girls at single-sex schools showed stronger preferences for low-pitched voices than girls at mixed sex schools, a difference that was particularly evident amongst the older girls. Further, there was a non-significant trend for the presence of brothers to covary positively with preference for higher pitch in both boys and girls. Together, these findings suggest that girls who have little experience of hearing younger boys' voices may base their perceptions of vocal attractiveness on older men (e.g. teachers or relatives) with lower-pitched voices.

In sum, although significant effects of school type and siblings were detected, the effects were not in line with predictions. Namely, children at mixed-sex schools were not clearly better at selecting those cues found attractive by adults to be more attractive compared with children at single-sex schools; and there was no moderating effect of opposite-sex siblings on the effects of school type. The pattern of results is not even wholly internally consistent between Year 1 and Year 2, suggesting that the relationship between experience and preference cannot be as robust and straightforward as predicted. In particular, there may be other differences between the pupils at the single-sex and mixed-sex schools that could influence their ratings, including cultural or environmental differences, some of which were reviewed in Chapter 1.

Long-term variation: general discussion

Section 3 set out to look at facultative changes over the long term. Chapters 6 and 7 examined how adolescence and puberty, and individual differences in experience around that time, could impact upon judgments of cues of mate quality found in the face and voice. It found good evidence that adolescents' judgments of facial attractiveness develop during adolescence in relation to female faces, and a little evidence that judgments may change similarly in relation to male faces. However, even the youngest groups of children, aged 11 – 13, demonstrated preferences for the more symmetric, more feminine and more average faces, implying that these supposed cues of genetic quality are also relevant to non-sexual contexts. The results also showed that girls' preferences for low pitched male voices increased with age, and that boys' preferences for high pitched female voices decreased with age. This contrasts with the demonstrated adult preference for high pitched female voices, and recommends further study on how attractiveness judgments continue to change into adulthood. A complex pattern of associations between pubertal development and preference was detected, suggesting that further work is required to unravel the links between puberty and attractiveness judgments. Chapter 7 found some modulatory effects of school type and the presence of brothers on attractiveness judgments, provoking the question of whether such school type or siblings effects on attractiveness judgments persist even longer term, into adulthood.

Chapter 8: Concluding remarks

The foregoing data chapters have presented studies that considered individual variation in human mate choice across three different time scales. The experimental findings were discussed at the end of each Chapter, and general implications of the topic were outlined in the discussion at the end of each Section. In this chapter, the results will be considered with reference to the initial premise set out in Chapter 1: that a general biological principle allowing for functional, facultative adjustment of human mate choice decisions might give rise to adaptive variation across short-, medium- and long-term time scales. This premise demanded that the research hypotheses, the experimental design, and the interpretation of the results, give consideration to biological function. Let us see how this played out in the data chapters, and how it reflects upon the initial premise.

Attractiveness judgment changes that take place across the short-term time scale were considered in Section 1 with reference to a chemical, androstadienone, that was found to influence women's judgments of men's attractiveness. As discussed in Chapter 1, the investigation of adjustment of mating behaviours over the short term, as a reaction to sensory modulators, is not novel, although the olfactory channel presents fascinating possibilities: it allows for the transmission of information which is potentially more direct and less filtered than channels such as vision, leading some to argue even that it may form the basis of the development of attraction judgments in the other senses (Kohl et al. 2001). The decision to test androstadienone for its effect on judgments of attractiveness was made on the basis of several strands of evidence that it performs some sort of sexual role: namely, research has found possibly sexually-dimorphic production rates beginning in puberty (see Gower & Ruparelia 1993 for an overview), sexually dimorphic and sex-orientation-specific responses (Savic et al. 2001; Savic et al. 2005; Boulkroune et al. 2007), production in the axilla where hair might assist in communicative dissemination of odours (Cohn 1994), and chemical similarity to sexual signals in other species (Signoret & du Mesnil du Buisson 1961). The assumption that modulation of attractiveness judgments is biologically relevant also gave rise to the recommendations for future work discussed at the end of Section 1, namely, that research should focus on whether androstadienone production corresponds to other aspects of men's mate quality.

The menstrual cycle was taken up as the progenitor of changes in mate choice behaviour over the medium term in Section 2. The theoretical position that an individual's attractiveness should be roughly equivalent in any modality motivated both experiments in Chapter 4. In the second study, women were rated more attractive at high-fertility, although the raters who had the most information about a woman's appearance were not better able to pick out the high-fertility recording as more attractive than the raters who had less information. As discussed at the end of Section 2, this could provide an insight into the functional beneficiary of these attractiveness changes: whether they constitute signals, cues, or coercion (Scott-Phillips 2008). Chapter 5 showed that masculinity and self-resemblance affected women's attractiveness judgments, and that these factors were influenced by relationship context. Again, the division of relationship types into the short-term and long-term reflects a functional understanding of reproductive strategies (Gangestad & Simpson 2000), and was productive in identifying a context-specific and hierarchical mate-choice decision rule.

The research hypothesis of Section 3 took as starting points the functional positions that judgments of attractiveness help us choose mates (see e.g. Rhodes 2006), and that psychological capabilities may come online as they are needed (Andersson 1994; Cartwright 2000). It investigated whether children were able to detect mate quality only as they went through puberty and reproductive choice became relevant, and whether individual differences in puberty predicted differences in attractiveness judgments. This functionally-motivated investigation gave rise to findings of age-related changes in judgments that were broadly in line with predictions.

In all of these cases, functional considerations have been essential to the formation or interpretation of the studies. As such, the research places itself within the discipline known as evolutionary psychology, whose functional motivations have been subject to criticism (see, e.g., Laland & Brown 2002 for a review). As emphasised in the summaries above, many of the functionally-motivated building blocks of the experimental work undertaken in the foregoing chapters are intangible: they are mentalistic, logically derived, or conceptual. That is, despite being biologically motivated, the studies work with abstract concepts such as genetic quality, masculinity, or even attractiveness, that are thought to represent constellations of more objective measurements. Of course the studies also employ specific, concrete elements; androstadienone is one example; but even then, as discussed in Section 1, a great deal is unclear about how androstadienone might be released and detected, and of course the work is motivated on the (abstract) assumption that androstadienone has an adaptive significance. This reliance on the abstract is no doubt a great source of frustration to those used to working with more quantifiable, objective objects or measurements. Yet this deficiency can also be used to inspire other research that attempts to pin down the biochemical pathways, or the neurological underpinnings, or the numerical evidence of increased reproductive success, that have been assumed by the behavioural research. Indeed, the school of evolutionary psychology, almost by definition, has made more efforts than other branches of psychology to link mentalistic concepts to biological knowledge; in the parlance of the field, this unites the proximate and the ultimate. Instead of dismissing evolutionary psychology's reliance on functional explanations, a more productive approach may be to look to integrate these functional explanations with the tangible foundations that are often associated more readily with other branches of science.

The approach of evolutionary psychology suggests that we should seek to understand human behaviour in the light of evolution and adaptation (c.f. Dobzhansky 1973). One thing we do not know is whether adaptive pressures should be used to explain only the most major swathes of our psychology, or also more minor variations. This thesis is in part a product of the Zeitgeist that attempts to apply adaptive reasoning even to individual variation (c.f. Wilson 1994; Wilson 1998; Nettle 2006). This topic of individual variation in human mate choice, then, falls across the boundary of what we currently understand about adaptations versus non-adaptive variation. Yet there is also an epistemological question at issue relating to the evolutionary psychology research paradigm. As is standard within the field, the experimental findings in the foregoing chapters have often been interpreted with reference to this explicative paradigm, even when they conflict with the original hypotheses. So for instance, the experiments of Section 3 started with the hypothesis that children would only be able to identify symmetry, averageness and sexual dimorphism as attractive when they became able to reproduce. The finding that the younger children also used these parameters in their judgments of attractiveness was used not to comment upon the original hypothesis, but instead was put forward as evidence that symmetry, averageness and sexual dimorphism must have non-sexual significance. This type of re-interpretation of the results to fit an adaptive framework has met with criticism of the evolutionary psychology research discipline on the basis that it provokes 'just so' stories. Yet one such just-so story must be correct, and on that basis alone, they should not be banned outright (Dennett 1995). Further, and more importantly, individual studies that take place within a scientific paradigm cannot themselves comment upon that paradigm (Kuhn 1962). Having said this, individual studies can form part of a body of evidence which shows whether that paradigm lends itself to the asking of useful questions, and the level of investigation (broad scale or minute variation) at which the paradigm may be used. The conventions of the evolutionary psychology research paradigm are that behaviour is frequently adaptive, and should be interpreted as such. The usefulness of evolutionary psychology in framing and provoking good questions will be judged by future historians of science with reference to the success or otherwise of experiments based within it, perhaps including, albeit as a minute part, the studies reported in this thesis.

Bibliography

Abitbol, J., de Brux, J., Millot, G., Masson, M.-F., Mimoun, O. L., Pau, H. & Abitbol, B. 1989. Does a hormonal vocal cord cycle exist in women? Study of vocal premenstrual syndrome in voice performers by videostroboscopy-glottography and cytology on 38 women. *Journal of Voice*, 3(2), 157-162.

Abitbol, J., Abitbol, P. & Abitbol, B. 1999. Sex hormones and the female voice. *Journal of Voice*, 13(3), 424-446.

Alexander, J. & Stimson, W. H. 1988. Sex hormones and the course of parasitic infection *Parasitology Today*, 4(7), S1-S2.

Alley, T. R. & Cunningham, M. R. 1991. Averaged faces are attractive, but very attractive faces are not average. *Psychological Science*, 2(2), 123-125.

Allison, D. B., Neale, M. C., Kezis, M. I., Alfonso, V. C., Heshka, S. & Heymsfield, S. B. 1996. Assortative mating for relative weight: genetic implications. *Behavior Genetics*, 26, 103-111.

Anastasi, J. S. & Rhodes, M. G. 2005. An own-age bias in face recognition for children and older adults. *Psychonomic Bulletin & Review*, 12(6), 1043-1047.

Andersson, M. 1994. Sexual Selection. Princeton, NJ: Princeton University Press.

Anon. 2005. So much more to know .... *Science magazine*, 309(5731), 78-102.

Apanius, V., Penn, D. J., Slev, P., Ruff, L. R. & Potts, W. K. 1997. The nature of selection on the major histocompatibility complex. *Critical Reviews in Immunology*, 17, 179-224.

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.

Apter, D. 1980. Serum steroids and pituitary hormones in female puberty: A partly longitudinal study. *Clinical Endocrinology*, 12(2), 107-120.

Apter, D. & Vihko, R. 1983. Early menarche, a risk factor for breast cancer, indicates early onset of ovulatory cycles. *Journal of Clinical Endocrinology and Metabolism*, 57(1), 82-86.

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

Arum, R., Roksa, J. & Budig, M. J. 2008. The romance of college attendance: Higher education stratification and mate selection. *Research in Social Stratification and Mobility*, 26(2), 107-121.

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 nonconception menstrual cycles of 32 normal women. *Human Reproduction*, 12, 2607-2613.

Baird, D. D., Weinberg, C. R., Zhou, H., Kamel, F., McConnaughey, D. R., Kesner, J. S. & Wilcox, A. J. 1999. Preimplantation urinary hormone profiles and the probability of conception in healthy women. *Fertility and Sterility*, 71, 40-49.

Bakker, T. C. M., Künzler, R. & Mazzi, D. 1999. Condition-related mate choice in sticklebacks. *Nature*, 401, 234.

Bateson, P. 1978. Sexual imprinting and optimal outbreeding. *Nature*, 273, 659 - 660.

Bateson, P. 1980. Optimal outbreeding and the development of sexual preferences in Japanese quail. *Zeitschrift fur Tierpsychologie*, 53(231-244).

Bateson, P. 1982. Preferences for cousins in Japanese quail. *Nature*, 295, 236 - 237.

Behr Davis, C. & Lee Davis, M. 1993. The effects of premenstrual syndrome (PMS) on the female singer. *Journal of Voice*, 7(4), 337-353.

Behre, H. M., Kuhlage, J., Ganer, C., Sonntag, B., Schem, C., Schneider, H. P. G. & Nieschlag, E. 2000. Prediction of ovulation by urinary hormone measurements with the home use ClearPlan(R) Fertility Monitor: comparison with transvaginal ultrasound scans and serum hormone measurements. *Human Reproduction*, 15(12), 2478-2482.

Belsky, J., Steinberg, L. & Draper, P. 1991. Childhood experience, interpersonal development, and reproductive strategy: An evolutionary theory of socialization. *Child Development*, 62, 647-670.

Belsky, J., Steinberg, L. D., Houts, R. M., Friedman, S. L., DeHart, G., Cauffman, E., Roisman, G. I., Halpern-Felsher, B. L. & Susman, E. 2007. Family rearing antecedents of pubertal timing. *Child Development*, 78(4), 1302-1321.

Ben-Ari, E. T. 1998. Pheromones: What's in a name? *Bioscience*, 48(7), 505-511.

Bensafi, M., Brown, W. M., Khan, R., Levenson, B. & Sobel, N. 2004a. Sniffing human sex-steroid derived compounds modulates mood, memory and autonomic nervous system function in specific behavioral contexts. *Behavioural Brain Research*, 152(1), 11-22.

Bensafi, M., Tsutsui, T., Khan, R., Levenson, R. W. & Sobel, N. 2004b. Sniffing a human sex-steroid derived compound affects mood and autonomic arousal in a dose-dependent manner. *Psychoneuroendocrinology*, 29, 1290-1299.

Benson, P. J. & Perrett, D. I. 1991. Synthesizing continuous-tone caricatures. *Image and Vision Computing*, 9, 123-129.

Benson, P. J. & Perrett, D. I. 1993. Extracting prototypical facial images from exemplars. *Perception*, 22, 257-262.

Benton, D. 1982. The influence of androstenol - a putative human pheromone - on mood throughout the menstrual cycle. *Biological Psychology*, 15(3-4), 249-256.

Benton, D. & Wastell, V. 1986. Effects of androstenol on human sexual arousal. *Biological Psychology*, 22(2), 141-147.

Bereczkei, T., Gyuris, P., Koves, P. & Bernath, L. 2002. Homogamy, genetic similarity, and imprinting; parental influence on mate choice preferences. *Personality And Individual Differences*, 33(5), 677-690.

Bereczkei, T., Gyuris, P. & Weisfeld, G. E. 2004. Sexual imprinting in human mate choice. *Proceedings of the Royal Society - Biological Sciences (Series B)*, 271(1544), 1129-1134.

Bernatchez, L. & Landry, C. 2003. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *Journal of Evolutionary Biology*, 16(3), 363-377.

Berscheid, E., Dion, K., Walster, E. & Walster, G. W. 1971. Physical attractiveness and dating choice: A test of the matching hypothesis. *Journal of Experimental Social Psychology*, 7(2), 173-189.

Bhatnagar, K. P. & Smith, T. D. 2001. The human vomeronasal organ. III. Postnatal development from infancy to the ninth decade. *Journal of Anatomy*, 199(289-302).

Bjorksten, T. A., Fowler, K. & Pomiankowski, A. 2000. What does sexual trait FA tell us about stress? *Trends In Ecology & Evolution*, 15(4), 163-166.

Black, S. L. & Biron, C. 1982. Androstenol as a human pheromone: no effect on perceived physical attractiveness. *Behavioral and Neural Biology*, 34, 326-330.

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

Boulkroune, N., Wang, L., March, A., Walker, N. & Jacob, T. J. C. 2007. Repetitive olfactory exposure to the biologically significant steroid androstadienone causes a hedonic shift and gender dimorphic changes in olfactory-evoked potentials. *Neuropsychopharmacology*, 32, 1822–1829.

Brase, G. L. & Walker, G. 2004. Male sexual strategies modify ratings of female models with specific waist-to-hip ratios. *Human Nature*, 15(2), 209-224.

Braude, S., Tang-Martinez, Z. & Taylor, G. T. 1999. Stress, testosterone, and the immunoredistribution hypothesis. *Behavioral Ecology*, 10(3), 345-350.

Bressan, P., Bertamini, M., Nalli, A. & Zanutto, A. 2008. Men do not have a stronger preference than women for self-resemblant child faces. *Archives of Sexual Behavior.*

Brooks-Gunn, J., Warren, M. P., Rosso, M. P. & Gargiulo, J. 1987. Validity of self-report measures of girls' pubertal status. *Child Development*, 58, 829-841.

Brooksbank, B. W. L. & Haslewood, G. A. D. 1961. The estimation of androst-16-en-3α-ol in human urine. Partial synthesis of androstenol and of its β-glucosiduronic acid. *Biochemical Journal*, 80, 488-496.

Brown, J. L. & Eklund, A. 1994. Kin recognition and the Major Histocompatibility Complex: An integrative review. *American Naturalist*, 143(3), 435-461.

Brown, J. L. 1997. A theory of mate choice based on heterozygosity. *Behavioral Ecology*, 8(1), 60-65.

Brown, J. L. 1999. The new heterozygosity theory of mate choice and the MHC. *Genetica*, 104, 215-221.

Brown, T. A., Cash, T. F. & Noles, S. W. 1986. Perceptions of physical attractiveness among college students: selected determinants and methodological matters. *The Journal of Social Psychology*, 126(3), 305-316.

Brown, W. M., Cronk, L., Grochow, K., Jacobson, A., Liu, C. K., Popović, Z. & Trivers, R. L. 2005. Dance reveals symmetry especially in young men. *Nature*, 438(7071), 1148-1150.

Buckingham, G., DeBruine, L. M., Little, A. C., Welling, L. L. M., Conway, C. A., Tiddeman, B. P. & Jones, B. C. 2006. Visual adaptation to masculine and feminine faces influences generalized preferences and perceptions of trustworthiness. *Evolution and Human Behavior*, 27(5), 381-389.

Burgess, A. 1990. Co-education - the disadvantages for schoolgirls. *Gender and Education*, 2(1), 91-95.

Burley, N. 1979. The evolution of concealed ovulation. *The American Naturalist*, 114(6), 835-858.

Burley, N. T. & Foster, V. S. 2006. Variation in female choice of mates: condition influences selectivity. *Animal Behaviour*, 72(3), 713-719.

Burnham, T. C., Chapman, J. F., Gray, P. B., McIntyre, M. H., Lipson, S. F. & Ellison, P. T. 2003. Men in committed, romantic relationships have lower testosterone. *Hormones and Behavior*, 44(2), 119-122.

Burriss, R. P. & Little, A. C. 2006. Effects of partner conception risk phase on male perception of dominance in faces. *Evolution and Human Behavior*, 27(4), 297-305.

Burt, D. M. & Perrett, D. I. 1997. Perceptual asymmetries in judgements of facial attractiveness, age, gender, speech and expression. *Neuropsychologia*, 35(5), 685-693.

Buss, D. M. & Schmitt, D. P. 1993. Sexual Strategies Theory: An evolutionary perspective on human mating. *Psychonomic Review*, 100(2), 204-232.

Buss, D. M. 1994. The evolution of desire: strategies of human mating. New York: Basic Books.

Buston, P. M. & Emlen, S. T. 2003. Cognitive processes underlying human mate choice: The relationship between self-perception and mate preference in Western society. *Proceedings of the National Academy of Sciences of the United States of America*, 100(15), 8805-8810.

Butler, S., Gilchrist, I. D., Burt, D. M., Perrett, D. I., Jones, E. & Harvey, M. 2005. Are the perceptual biases found in chimeric face processing reflected in eye-movement patterns? *Neuropsychologia*, 43(1), 52-59.

Candolin, U. 2003. The use of multiple cues in mate choice. *Biological Reviews of the Cambridge Philosophical Society*, 78(4), 575-595.

Cardenas, R. A. & Harris, L. J. 2007. Do women's preferences for symmetry change across the menstrual cycle? *Evolution and Human Behavior*, 28(2), 96-105.

Carskadon, M. A. & Acebo, C. 1993. A self-administered rating scale for pubertal development. *Journal of Adolescent Health*, 14(3), 190-195.

Cartwright, J. 2000. Evolution and Human Behavior: Darwinian Perspectives on Human Nature. Cambridge, MA: MIT Press.

Castiello, U., Zucco, G. M., Parma, V., Ansuini, C. & Tirindelli, R. 2006. Cross-modal interactions between olfaction and vision when grasping. *Chemical Senses*, 31(7), 665-671.

Cavior, N. & Lombardi, D. A. 1973. Developmental aspects of judgment of physical attractiveness in children. *Developmental Psychology*, 8(1), 67-71.

Chae, S. W., Chol, G., Kang, H. J., Choi, J. O. & Jin, S. M. 2001. Clinical analysis of voice change as a parameter of premenstrual syndrome. *Journal of Voice*, 15(2), 278-283.

Chaieb, K., Hajlaoui, H., Zmantar, T., Kahla-Nakbi, A. B., Rouabhia, M., Mahdouani, K. & Bakhrouf, A. 2007. The chemical composition and biological activity of clove essential oil, *Eugenia caryophyllata* (*Syzigium aromaticum* L. Myrtaceae): a short review. *Phytotherapy Research*, 21(6), 501-506.

Chopra, A., Baur, A. & Hummel, T. 2008. Thresholds and chemosensory event-related potentials to malodors before, during, and after puberty: Differences related to sex and age. *Neuroimage*, 40(3), 1257-1263.

Chu, S. & Downes, J. J. 2000. Odour-evoked autobiographical memories: psychological investigations of Proustian phenomena. *Chemical Senses*, 25(1), 111-116.

Chu, S., Hardaker, R. & Lycett, J. E. 2007. Too good to be `true'? The handicap of high socio-economic status in attractive males. *Personality And Individual Differences*, 42(7), 1291.

Cleveland, W. W. & Savard, K. 1964. Studies of excretion of androst-16-en-3α-ol. *Journal of Clinical Endocrinology and Metabolism*, 24, 983-987.

Cohn, B. A. 1994. In search of human skin pheromones. *Archives of Dermatology*, 130(8), 1048-1051.

Cohn, L. D. & Adler, N. E. 1992. Female and male perceptions of ideal body shapes: Distorted views among caucasian college students. *Psychology of Women Quarterly*, 16(1), 69-79.

Colditz, G. A., Hankinson, S. E., Hunter, D. J., Willett, W. C., Manson, J. E., Stampfer, M. J., Hennekens, C., Rosner, B. & Speizer, F. E. 1995. The use of estrogens and progestins and the risk of breast cancer in postmenopausal women. *The New England Journal of Medicine*, 332, 1589-1593.

Colegrave, N., Kotiaho, J. S. & Tomkins, J. L. 2002. Mate choice or polyandry: reconciling genetic compatibility and good genes sexual selection. *Evolutionary Ecology Research*, 4, 911-917.

Collins, S. A. 2000. Men's voices and women's choices. *Animal Behaviour*, 60, 773-780.

Collins, S. A. & Missing, C. 2003. Vocal and visual attractiveness are related in women. *Animal Behaviour*, 65(5), 997-1004.

Connolly, J. M., Slaughter, V. & Mealey, L. 2004. The development of preferences for specific body shapes. *Journal of Sex Research*, 41(1), 5-15.

Cooke, F. 1976. How do we choose our mate? *New Scientist*, 81, 325-327.

Cooke, F. 1978. Early learning and its effect on population structure. Studies of a wild population of snow geese. *Zeitschrift fur Tierpsychologie*, 344-358.

Cooper, P. A., Geldart, S. S., Mondloch, C. J. & Maurer, D. 2006. Developmental changes in perceptions of attractiveness: a role of experience? *Developmental Science*, 9(5), 530-543.

Cornwell, R. E., Palmer, C. T. & Davis, H. P. 2002. More women (and men) that never evolved. *Behavioral and Brain Sciences*, 23, 598-599.

Cornwell, R. E., Boothroyd, L., Burt, D. M., Feinberg, D. R., Jones, B. C., Little, A. C., Pitman, R., Whiten, S. & Perrett, D. I. 2004. Concordant preferences for opposite-sex signals? Human pheromones and facial characteristics. *Proceedings of the Royal Society - Biological Sciences (Series B)*, 271(1539), 635-640.

Cornwell, R. E., Law Smith, M. J., Boothroyd, L. G., Moore, F. R., Davis, H. P., Stirrat, M., Tiddeman, B. & Perrett, D. I. 2006. Reproductive strategy, sexual development and attraction to facial characteristics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1476), 2143–2154.

Costabile, K. A. & Klein, S. B. 2005. Finishing strong: Recency effects in juror judgments. *Basic and Applied Social Psychology*, 27(1), 47 - 58.

Cowley, J. J., Johnson, A. T. & Brooksbank, B. W. L. 1977. The effect of two odorous compounds on performance in an assessment-of-people test. *Psychoneuroendocrinology*, 2, 159-172.

Cowley, J. J., Harvey, F., Johnson, A. T. & Brooksbank, B. W. L. 1980. Irritability and depression during the menstrual cycle—possible role for exogenous pheromone. *Irish Journal of Psychology*, 3, 143-156.

Cowley, J. J. & Brooksbank, B. W. L. 1991. Human exposure to putative pheromones and changes in aspects of social behaviour. *Journal of Steroid Biochemistry and Molecular Biology*, 39(4B), 647-659.

Cross, J. F. & Cross, J. 1971. Age, sex, race and the perception of facial beauty. *Developmental Psychology*, 5(3), 433-439.

Cutler, G. B., Jr., Schiebinger, R. J., Albertson, B. D., Cassorla, F. G., Chrousos, G. P. & Comite, F. 1990. Adrenarche (human and animal). In: *Control of the Onset of Puberty* (Ed. by Grumbach, M. M., Sizonenko, P. C. & Aubert, M. L.), pp. 506-533. Baltimore: Williams & Wilkins.

Dabbs, J. M. 1997. Testosterone, smiling, and facial appearance. *Journal of Nonverbal Behavior*, 21(1), 45-55.

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

Dahl, R. E. 2004. Adolescent brain development: Vulnerabilities and opportunities. In: *Annals of the New York Academy of Sciences* (Ed. by Dahl, R. E. & Spear, L. P.). New York.

Danel, D. & Pawlowski, B. 2006. Attractiveness of men's faces in relation to women's phase of menstrual cycle. *Collegium Antropologicum*, 30(2), 285-289.

Darwin, C. 1871. The Descent of Man, and Selection in Relation to Sex. London: John Murray.

David, A. S. 1993. Spatial and selective attention in the cerebral hemispheres in depression, mania, and schizophrenia. *Brain and Cognition*, 23(2), 166-180.

Davis, J. W. F. & O'Donald, P. 1976. Estimation of assortative mating preferences in the arctic skua. *Heredity*, 36, 235-244.

Davison, K. K. & Susman, E. J. 2001. Are hormone levels and cognitive ability related during early adolescence? *International Society for the Study of Behavioral Development*, 25(5), 416-428.

Dawkins, R. 1976. The Selfish Gene. Oxford: Oxford University Press.

Dawkins, R. & Krebs, J. R. 1979. Arms races between and within species. *Proceedings of the Royal Society of London Series B*, 205, 489-511.

de Heering, A. & Rossion, B. 2008. Prolonged visual experience in adulthood modulates holistic face perception. *PLoS ONE*, 3(5), e2317.

De Souza, M. J., Miller, B. E., Loucks, A. B., Luciano, A. A., Pescatello, L. S., Campbell, C. G. & Lasley, B. L. 1998. High frequency of luteal phase deficiency and anovulation in recreational women runners: Blunted elevation in Follicle-Stimulating Hormone observed during luteal-follicular transition. *Journal of Clinical Endocrinology and Metabolism*, 83(12), 4220-4232.

DeBruine, L. M. 2002. Facial resemblance enhances trust. *Proceedings of the Royal Society of London Series B*, 269, 1307-1312.

DeBruine, L. M. 2004. Facial resemblance increases the attractiveness of same-sex faces more than other-sex faces. *Proceedings Of The Royal Society Of London Series B-Biological Sciences*, 271(1551), 2085-2090.

DeBruine, L. M. 2005. Trustworthy but not lust-worthy: context-specific effects of facial resemblance. *Proceedings of the Royal Society - Biological Sciences (Series B)*, 272(1566), 919-922.

DeBruine, L. M., Jones, B. C. & Perrett, D. I. 2005. Women's attractiveness judgments of self-resembling faces change across the menstrual cycle. *Hormones and Behavior*, 47(4), 379.

DeBruine, L. M., Jones, B. C., Little, A. C., Boothroyd, L. G., Perrett, D. I., Penton-Voak, I. S., Cooper, P. A., Penke, L., Feinberg, D. R. & Tiddeman, B. P. 2006. Correlated preferences for facial masculinity and ideal or actual partner's masculinity.

Proceedings Of The Royal Society Of London Series B-Biological Sciences, 273(1592), 1355-1360.

DeBruine, L. M., Jones, B. C., Little, A. C. & Perrett, D. I. 2008. Social perception of facial resemblance in humans. *Archives of Sexual Behavior*, 37(1), 64-77.

Decreau, R. A., Marson, C. M., Smith, K. E. & Behan, J. M. 2003. Production of malodorous steroids from androsta-5,16-dienes and androsta-4,16-dienes by Corynebacteria and other human axillary bacteria. *The Journal of Steroid Biochemistry and Molecular Biology*, 87(4-5), 327-336.

Demattè, M. L., Österbauer, R. & Spence, C. 2007. Olfactory cues modulate facial attractiveness. *Chemical Senses*, 32(6), 603-610.

Dennett, D. 1995. Darwin's Dangerous Idea: Evolution and the Meanings of Life. New York: Simon & Schuster.

DeWitt, T. J., Sih, A. & Wilson, D. S. 1998. Costs and limits of phenotypic plasticity. *Trends In Ecology & Evolution*, 13(2), 77-81.

Dion, K. K. 1973. Young children's stereotyping of facial attractiveness. *Developmental Psychology*, 9(2), 183-188.

Dobzhansky, T. 1973. Nothing in biology makes sense except in the light of evolution. *American Biology Teacher*, 35, 125-129.

Doherty, P. C. & Zinkernagel, R. M. 1975. Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature*, 256, 50-52.

Domb, L. G. & Pagel, M. 2001. Sexual swellings advertise female quality in wild baboons. *Nature*, 410(6825), 204-206.

Dorn, L. D., Dahl, R. E., Woodward, H. R. & Biro, F. 2006. Defining the boundaries of early adolescence: A user's guide to assessing pubertal status and pubertal timing in research with adolescents. *Applied Developmental Science*, 10(1), 30-56.

Dornbusch, S. M., Carlsmith, J. M., Gross, R. T., Martin, J. A., Jennings, D., Rosenberg, A. & Duke, P. 1981. Sexual development, age, and dating: A comparison of biological and social influences upon one set of behaviors. *Child Development*, 52(1), 179-185.

Dorries, K. M., Adkins Regan, E. & Halpern, B. P. 1997. Sensitivity and behavioural responses to the pheromone androstenone are not mediated by the vomeronasal organ in domestic pigs. *Brain, Behavior and Evolution*, 49(1), 53-62.

Doty, R. L., Ford, M., Preti, G. & Huggins, G. R. 1975. Changes in the intensity and pleasantness of human vaginal odors during the menstrual cycle. *Science*, 190(4221), 1316-1318.

Duke, P. M., Litt, I. F. & Gross, R. T. 1980. Adolescents' self-assessment of sexual maturation. *Pediatrics*, 66(6), 918-920.

Eagly, A. H. & Wood, W. 1985. Gender and influenceability: Stereotype versus behavior. In: *Women, gender, and social psychology* (Ed. by O'Leary, V. E., Unger, R. K. & Wallston, B. S.), pp. 225-256. Hillsdale, NJ: Erlbaum.

Eagly, A. H. 1987. Sex Differences in Social Behavior. Hillsdale, NJ: Lawrence Erlbaum.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G. & Longo, L. C. 1991. What is beautiful is good: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, 110, 109-128.

Ehrhardt, A. A. & Meyer-Bahlburg, H. F. 1994. Psychosocial aspects of precocious puberty. *Hormone Research*, 41(supplement 2), 30-35.

Ekman, P., Brattesani, K. A., O'Sullivan, M. & Friesen, W. V. 1979. Does image size affect judgments of the face? *Journal of Nonverbal Behavior*, 4(1), 57-61.

Ekman, P., Friesen, W. V., O'Sullivan, M. & Scherer, K. 1980. Relative importance of face, body, and speech in judgments of personality and affect. *Journal of Personality and Social Psychology*, 38(2), 270-277.

Enlow, D. H. 1990. Facial Growth, 3rd edn. Philadelphia, USA: W.B. Saunders Company.

Enlow, D. H. & Hans, M. G. 1996. Essentials of Facial Growth. Philadelphia, PA: Saunders.

Epstein, J., Klinkenberg, W. D., Wiley, D. & McKinley, L. 2001. Insuring sample equivalence across internet and paper-and-pencil assessments. *Computers in Human Behavior*, 17(3), 339-346.

Evans, S., Neave, N., Wakelin, D. & Hamilton, C. 2008. The relationship between testosterone and vocal frequencies in human males. *Physiology and Behavior*, 93(4-5), 783-788.

Feinberg, D. R., Jones, B. C., DeBruine, L. M., Moore, F. R., Law Smith, M. J., Cornwell, R. E., Tiddeman, B. P., Boothroyd, L. G. & Perrett, D. I. 2005a. The voice and face of woman: One ornament that signals quality? *Evolution and Human Behavior*, 26(5), 398-408.

Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M. & Perrett, D. I. 2005b. Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour*, 69, 561-568.

Feinberg, D. R., Jones, B. C., Law Smith, M. J., Moore, F. R., DeBruine, L. M., Cornwell, R. E., Hillier, S. G. & Perrett, D. I. 2006. Menstrual cycle, trait estrogen level, and masculinity preferences in the human voice. *Hormones and Behavior*, 49(2), 215-222.

Feinberg, D. R., DeBruine, L. M., Jones, B. C. & Little, A. C. 2008. Correlated preferences for men's facial and vocal masculinity. *Evolution and Human Behavior*, 29(4), 233-241.

Fernald, A. & Kuhl, P. 1987. Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, 10(3), 279-293.

Field, A. 2005. Discovering Statistics Using SPSS, 2nd edn. London: SAGE Publications Ltd.

Filsinger, E. E., Braun, J. J., Monte, W. C. & Linder, D. E. 1984. Human (Homo sapiens) responses to the pig (Sus scrofa) sex pheromone 5 alpha-androst-16-en-3-one. *Journal of Comparative Psychology*, 98(2), 219-222.

Filsinger, E. E., Braun, J. J. & Monte, W. C. 1985. An examination of the effects of putative pheromones on human judgments. *Ethology and Sociobiology*, 6, 227-236.

Fink, B. & Penton-Voak, I. 2002. Evolutionary psychology of facial attractiveness. *Current Directions In Psychological Science*, 11(5), 154-158.

Fink, B., Neave, N., Manning, J. T. & Grammer, K. 2005. Facial symmetry and the 'big-five' personality factors. *Personality And Individual Differences*, 39(3), 523-529.

Fink, B., Grammer, K. & Matts, P. J. 2006. Visible skin color distribution plays a major role in the perception of age, attractiveness and health in female faces. *Evolution and Human Behavior*, 27(6), 433-442.

Finkel, E. J., Eastwick, P. W. & Matthews, J. 2007. Speed-dating as an invaluable tool for studying romantic attraction: A methodological primer. *Personal Relationships*, 14(1), 149-166.

Fisher, M. L. 2004. Female intrasexual competition decreases female facial attractiveness. *Proceedings of the Royal Society B: Biological Sciences*, 271(0), S283-S285.

Fitch, W. T. & Jay, G. 1999. Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 106(3), 1511-1522.

Fluhmann, C. F. 1957. Irregularities of the stages of the menstrual cycle. *Western Journal of Surgery, Obstetrics, and Gynecology*, 65, 265-277.

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

Forest, M. G. 1989. Physiological changes in circulating androgens. *Pediatric and Adolescent Endocrinology*, 19, 104-129.

Frost, P. 1994. Preferences for darker faces in photographs at different phases of the menstrual cycle: preliminary assessment of evidence for a hormonal relationship. *Perceptual and Motor Skills*, 79, 507-514.

Fukushima, S., Akane, A., Matsubara, K., Shiono, H., Morishita, H. & Nakada, F. 1991. Simultaneous determination of testosterone and androstadienone (sex attractant) in human plasma by gas chromatography--mass spectrometry with high-resolution selected-ion monitoring. *Journal of Chromatography: Biomedical Applications*, 565(1-2), 35-44.

Gangestad, S. W. & Buss, D. M. 1993. Pathogen prevalence and human mate preferences. *Ethology and Sociobiology*, 14, 89-96.

Gangestad, S. W. & Thornhill, R. 1998. Menstrual cycle variation in women's preferences for the scent of symmetrical men.

Proceedings Of The Royal Society Of London Series B-Biological Sciences, 265(1399), 927-933.

Gangestad, S. W. & Simpson, J. A. 2000. The evolution of human mating: Trade-offs and strategic pluralism. *Behavioral and Brain Sciences*, 23, 573-644.

Gangestad, S. W., Simpson, J. A., Cousins, A. J., Garver-Apgar, C. E. & Niels Christensen, P. 2004. Women's preferences for male behavioral displays change across the menstrual cycle. *Psychological Science*, 15(3), 203-207.

Gangestad, S. W. & Thornhill, R. 2008. Human oestrus. *Proceedings of the Royal Society B*, 275(1638), 991-1000.

Garcia-Velasco, J. & Mondragon, M. 1991. The incidence of the vomeronasal organ in 1000 human subjects and its possible clinical significance. *Journal of Steroid Biochemistry and Molecular Biology*, 39, 561-563.

Garcia, J. E., Jones, G. S. & Wright Jr, G. L. 1981. Prediction of the time of ovulation. *Fertility and Sterility*, 36(3), 308-315.

Garver-Apgar, C. E., Gangestad, S. W. & Thornhill, R. 2008. Hormonal correlates of women's mid-cycle preference for the scent of symmetry. *Evolution and Human Behavior*, 29(4), 223-232.

Getty, T. 2002. Signalling health versus parasites. *American Naturalist*, 159, 363-371.

Giersing, M., Lundstrom, K. & Andersson, A. 2000. Social effects and boar taint: significance for production of slaughter boars (Sus scrofa). *J. Anim Sci.*, 78(2), 296-305.

Giorgi, A., Weatherby, R. P. & Murphy, P. W. 1999. Muscular strength, body composition and health responses to the use of testosterone enanthate: a double blind study. *Journal of Science and Medicine in Sport*, 2(4), 341-355.

Goren, C. C., Sarty, M. & Wu, P. Y. K. 1975. Visual following and pattern discrimination of face like stimuli by newborn infants. *Pediatrics*, 56(4), 544-549.

Gosling, L. M. & Roberts, S. C. 2001. Scent-marking by male mammals: Cheat-proof signals to competitors and mates. *Advances in the Study of Behavior*, 30, 169-217.

Gower, D. B., Nixon, A. & Mallet, A. I. 1988. The significance of odorous steroids in axillary odour. In: *Perfumery: The psychology and biology of fragrance* (Ed. by Toller, S. V. & Dodd, G. H.), pp. 47-76. London: Chapman & Hall.

Gower, D. B. & Ruparelia, B. A. 1993. Olfaction in humans with special reference to odorous 16-androstenes: their occurrence, perception and possible social, psychological and sexual impact. *Journal of Endocrinology*, 137(2), 167-187.

Gower, D. B., Holland, K. T., Mallet, A. I., Rennie, P. J. & Watkins, W. J. 1994. Comparison of 16-Androstene steroid concentrations in sterile apocrine sweat and axillary secretions: Interconversions of 16-androstenes by the axillary microflora--a mechanism for axillary odour production in man? *Journal of Steroid Biochemistry and Molecular Biology*, 48(4), 409-418.

Grammer, K. & Thornhill, R. 1994. Human (Homo sapiens) facial attractiveness and sexual selection: The role of symmetry and averageness. *Journal of Comparative Psychology*, 108(3), 233-242.

Grammer, K., Renninger, L. & Fischer, B. 2004. Disco clothing, female sexual motivation, and relationship status: is she dressed to impress? *Journal of Sex Research*, 41(1), 66-74.

Gray, P. B., Kahlenberg, S. M., Barrett, E. S., Lipson, S. F. & Ellison, P. T. 2002. Marriage and fatherhood are associated with lower testosterone in males. *Evolution and Human Behavior*, 23(3), 193-201.

Gray, P. B., Chapman, J. F., Burnham, T. C., McIntyre, M. H., Lipson, S. F. & Ellison, P. T. 2004. Human male pair bonding and testosterone. *Human Nature*, 15(2), 119-131.

Griffiths, R. W. & Kunz, P. R. 1973. Assortative mating: a study of physiognomic homogamy. *Social Biology*, 20(4), 448-453.

Gross, M. R. 1996. Alternative reproductive strategies and tactics: diversity within sexes. *Trends In Ecology & Evolution*, 11(2), 92.

Grosser, B. I., Monti-Bloch, L., Jennings-White, C. & Berliner, D. L. 2000. Behavioral and electrophysiological effects of androstadienone, a human pheromone. *Psychoneuroendocrinology*, 25(3), 289-299.

Grossman, C. J. 1984. Regulation of the immune system by sex steroids. *Endocrinological Reviews*, 5(8), 435-455.

Grossman, C. J. 1985. Interactions between the gonadal steroids and the immune system. *Science*, 277(4684), 257-261.

Grumbach, M. M. & Styne, D. M. 1992. Puberty: Ontogeny, neuroendocrinology, physiology, and disorders. In: *Williams Textbook of Endocrinology* (Ed. by Wilson, J. D. & Foster, P. W.), pp. 1139-1231. Philadelphia: Saunders.

Grumbach, M. M. 2002. The neuroendocrinology of human puberty revisited. *Hormone Research*, 57(Suppl 2), 2-14.

Grumbach, M. M. & Styne, D. M. 2003. Puberty: Ontogeny, neuroendocrinology, physiology, and disorders. In: *Williams Textbook of Endocrinology* (Ed. by Larsen, P. R., Kronenberg, H. M., Melmed, S. & Polonsky, K. S.), pp. 1115-1286.

Gudgeon, K., Leader, L. & Howard, B. 1990. Evaluation of the accuracy of the home ovulation detection kit, Clearplan, at predicting ovulation. *The Medical Journal of Australia*, 152(7), 344 - 349.

Guida, M., Tommaselli, G. A., Palomba, S., Pellicano, M., Moccia, G., Di Carlo, C. & Nappi, C. 1999. Efficacy of methods for determining ovulation in a natural family planning program. *Fertility and Sterility*, 72(5), 900-904.

Gulyas, B., Keri, S., O'Sullivan, B. T., Decety, J. & Roland, P. E. 2004. The putative pheromone androstadienone activates cortical fields in the human brain related to social cognition. *Neurochemistry International*, 44(8), 595-600.

Gustavson, A. R., Dawson, M. E. & Bonett, D. G. 1987. Androstenol, a putative human pheromone, affects human (Homo sapiens) male choice performance. *Journal of Comparative Psychology*, 101(2), 210-212.

Haag, P. 1998. Single-sex education in grades K-12: What does the research tell us? In: *Separated by sex: A critical look at single-sex education for girls* (Ed. by Haag, P.). Washington, DC: American Association of University Women Educational Foundation.

Halberstadt, J. & Rhodes, G. 2000. The attractiveness of nonface averages: Implications for an evolutionary explanation of the attractiveness of average faces. *Psychological Science*, 11(4), 285-289.

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), 436-447.

Hamilton, B. L. 1977. An empirical investigation of the effects of heterogeneous regression slopes in analysis of covariance. *Educational and Psychological Measurement*, 37(3), 701-712.

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

Hansen, L., Bangsbo, J., Twisk, J. & Klausen, K. 1999. Development of muscle strength in relation to training level and testosterone in young male soccer players. *Journal of Applied Physiology*, 87(3), 1141-1147.

Haselton, M. G. & Miller, G. F. 2006. Women's fertility across the cycle increases the short-term attractiveness of creative intelligence. *Human Nature*, 17(1), 50-73.

Haselton, M. G., Mortezaie, M., Pillsworth, E. G., Bleske-Rechek, A. & Frederick, D. A. 2007. Ovulatory shifts in human female ornamentation: Near ovulation, women dress to impress. *Hormones and Behavior*, 51(1), 40-45.

Havlicek, J., Roberts, S. C. & Flegr, J. 2005. Women's preference for dominant male odour: effects of menstrual cycle and relationship status. *Biology Letters*, 1(3), 256-259.

Havlicek, J., Dvorakova, R., Bartos, L. & Flegr, J. 2006. Non-advertized does not mean concealed: Body odour changes across the human menstrual cycle. *Ethology*, 112(1), 81-90.

Havlicek, J., Saxton, T. K., Roberts, S. C., Jozifkova, E., Lhota, S., Valentova, J. & Flegr, J. 2008. He sees, she smells? Male and female reports of sensory reliance in mate choice and non-mate choice contexts. *Personality And Individual Differences*, 45, 564-569.

Havlicek, J. & Saxton, T. K. in press. The effect of diet on human bodily odors. In: *New Research on Food Habits* (Ed. by Hasegawa, K. & Takahashi, H.). New York: Nova Publishing.

Hays, W. S. T. 2003. Human pheromones: have they been demonstrated? *Behavioral Ecology and Sociobiology*, 54, 89-97.

Hedrick, P. W. & Black, F. L. 1997. HLA and mate selection: no evidence in South Amerindians. *American Journal of Human Genetics*, 61, 505-511.

Herz, R. S. & Cahill, E. D. 1997. Differential use of sensory information in sexual behavior as a function of gender. *Human Nature*, 8(3), 275-286.

Herz, R. S. & Inzlicht, M. 2002. Sex differences in response to physical and social factors involved in human mate selection - The importance of smell for women. *Evolution and Human Behavior*, 23(5), 359-364.

Higgins, M. B. & Saxman, J. H. 1989. Variations in vocal frequency perturbation across the menstrual cycle. *Journal of Voice*, 3(3), 233-243.

Hill, S. E. & Reeve, H. K. 2004. Mating games: the evolution of human mating transactions. *Behavioral Ecology*, 15(5), 748-756.

Hinsz, V. B. 1989. Facial resemblance in engaged and married couples. *Journal of Social and Personal Relationships*, 6(2), 223-229.

Holland, R. W., Hendriks, M. & Aarts, H. 2005. Smells like clean spirit. Nonconscious effects of scent on cognition and behavior. *Psychological Science*, 16(9), 689-693.

Hollien, H. & Malcik, E. 1967. Evaluation of cross-sectional studies of adolescent voice change in males. *Speech Monographs*, 34, 80-84.

Hopwood, D., Farrar, M. D., Bojar, R. A. & Holland, K. T. 2005. Microbial colonization dynamics of the axillae of an individual over an extended period. *Acta Dermato-Venereologica*, 85(4), 363-364.

Hopwood, N., Kelch, R. P., Hale, P. M., Mendes, T. M., Foster, C. M. & Beitins, I. Z. 1990. The onset of human puberty: Biological and environmental factors. In: *Adolescence and Puberty*

(Ed. by Bancroft, J. & Reinisch, J. M.). Oxford: Oxford University Press.

Hubert, G. D. & Carson, S. A. 1990. Adrenarche. *Adolescent and Pediatric Gynecology*, 3, 3-14.

Hughes, S. M., Harrison, M. A. & Gallup, J. G. G. 2002. The sound of symmetry: Voice as a marker of developmental instability. *Evolution and Human Behavior*, 23(3), 173-180.

Ibanez, L., de Zegher, F. & Potau, N. 1999. Anovulation after precocious pubarche: Early markers and time course in adolescence. *Journal of Clinical Endocrinology and Metabolism*, 84(8), 2691-2695.

Jackson, C. 2002. Can single-sex classes in co-educational schools enhance the learning experiences of girls and/or boys? An exploration of pupils' perceptions. *British Educational Research Journal*, 28(1), 37-48.

Jackson, L. A. & McGill, O. D. 1996. Body type preferences and body characteristics associated with attractive and unattractive bodies by african Americans and Anglo Americans. *Sex Roles*, 35(5), 295-307.

Jacob, S. & McClintock, M. K. 2000. Psychological state and mood effects of steroidal chemosignals in women and men *Hormones and Behavior*, 37, 57-78.

Jacob, S., Hayreh, D. J. S. & McClintock, M. K. 2001a. Context-dependent effects of steroid chemosignals on human physiology and mood. *Physiology and Behavior*, 74(1-2), 15-27.

Jacob, S., Kinnunen, L. H., Metz, J., Cooper, M. & McClintock, M. K. 2001b. Sustained human chemosignal unconsciously alters brain function. *Neuroreport*, 12(11), 2391-2394.

Jacob, S., Garcia, S., Hayreh, D. & McClintock, M. K. 2002a. Psychological effects of musky compounds: Comparison of androstadienone with androstenol and muscone. *Hormones and Behavior*, 42(3), 274-283.

Jacob, S., McClintock, M. K., Zelano, B. & Ober, C. 2002b. Paternally inherited HLA alleles are associated with women's choice of male odor. *Nature Genetics*, 30(2), 175-179.

Jaffe, K. & Chacon-Puignau, G. 1995. Assortative mating: sex differences in mate selection for married and unmarried couples. *Human Biology*, 67(1), 111-120.

James, W. 1890. Psychology: The briefer course: Henry Holt and Company.

Jasienska, 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 Series B-Biological Sciences*, 271(1545), 1213 - 1217.

Jessica, E. H., Elaine, T. S., Gina, M. C., Theresa, A. A. & Kenneth, J. 1999. Formants of children, women, and men: The effects of vocal intensity variation. *The Journal of the Acoustical Society of America*, 106(3), 1532-1542.

Jimenez, E. & Lockheed, M. E. 1989. Enhancing girls' learning through single-sex education: Evidence and a policy conundrum. *Educational Evaluation and Policy Analysis*, 11(2), 117-142.

Jin, K., Ho, H. N., Speed, T. P. & Gill, T. J. I. 1995. Reproductive failure and the major histocompatibility complex. *American Journal of Human Genetics*, 56, 1456-1467.

Johns, M. A., Feder, H. H., Komisaruk, B. R. & Mayer, A. D. 1978. Urine-induced ovulation in anovulatory rats may be a vomeronasal effect. *Nature*, 272, 446-448.

Johnson, M. H., Dziurawiec, S., Ellis, H. & Morton, J. 1991. Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, 40, 1-19.

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.

Johnstone, R. A. 1996. Multiple displays in animal communication:`backup signals' and `multiple messages'. *Philosophical Transactions: Biological Sciences*, 351(1337), 329-338.

Johnstone, R. A. 1997. The tactics of mutual mate choice and competitive search. *Behavioral Ecology and Sociobiology*, 40(1), 51.

Jones, B. C., Little, A. C., Burt, D. M. & Perrett, D. I. 2004a. When facial attractiveness is only skin deep. *Perception*, 33(5), 569-576.

Jones, B. C., Little, A. C., Feinberg, D. R., Penton-Voak, I. S., Tiddeman, B. P. & Perrett, D. I. 2004b. The relationship between shape symmetry and perceived skin condition in male facial attractiveness. *Evolution and Human Behavior*, 25(1), 24-30.

Jones, B. C., Little, A. C., Boothroyd, L., DeBruine, L. M., Feinberg, D. R., Smith, M. J. L., Cornwell, R. E., Moore, F. R. & Perrett, D. I. 2005a. Commitment to relationships and preferences for femininity and apparent health in faces are strongest on days of the menstrual cycle when progesterone level is high. *Hormones and Behavior*, 48(3), 283.

Jones, B. C., Little, A. C., Boothroyd, L., Feinberg, D. R., Cornwell, R. E., DeBruine, L. M., Roberts, S. C., Penton-Voak, I. S., Law Smith, M. J., Moore, F. R., Davis, H. P. & Perrett, D. I. 2005b. Women's physical and psychological condition independently predict their preference for apparent health in faces. *Evolution and Human Behavior*, 26, 451-457.

Jones, B. C., DeBruine, L. M. & Little, A. C. 2007. The role of symmetry in attraction to average faces. *Perception and Psychophysics*, 69(8), 1273-1277.

Jones, B. C., DeBruine, L. M., Perrett, D. I., Little, A. C., Feinberg, D. R. & Law Smith, M. J. 2008a. Effects of menstrual cycle phase on face preferences. *Archives of Sexual Behavior*, 37, 78-84.

Jones, B. C., Feinberg, D. R., DeBruine, L. M., Little, A. C. & Vukovic, J. 2008b. Integrating cues of social interest and voice pitch in men's preferences for women's voices. *Biology Letters*, 4(2), 192-194.

Jones, B. T., Jones, B. C., Thomas, A. P. & Piper, J. 2003. Alcohol consumption increases attractiveness ratings of opposite-sex faces: a possible third route to risky sex. *Addiction*, 98(8), 1069-1075.

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.

Kanwisher, N. & Yovel, G. 2006. The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1476), 2109–2128.

Karlson, P. & Lüscher, M. 1959. Pheromones: a new term for a class of biologically active substances. *Nature*, 183, 55-56.

Keller, A., Zhuang, H., Chi, Q., Vosshall, L. B. & Matsunami, H. 2007. Genetic variation in a human odorant receptor alters odour perception. *Nature*, 449(7161), 468.

Kendrick, K. M., Hinton, M. R. & Atkins, K. 1998. Mothers determine sexual preferences. *Nature*, 395, 229-230.

Kerin, J. 1982. Ovulation detection in the human. *Clinical Reproduction and Fertility*, 1(1), 27-54.
Keverne, E. B. 1983. Pheromonal influences on the endocrine regulation of reproduction. *Trends in Neurosciences*, 6, 381-384.
Kirk-Smith, M. D., Booth, D. A., Carroll, D. & Davies, P. 1978. Human social attitudes affected by androstenol. *Research Communications in Psychology, Psychiatry and Behavior*, 3(4), 379-384.
Kirk-Smith, M. D. & Booth, D. A. 1980. Effects of androstenone on choice of location in others' presence. In: *Olfaction and Taste VII* (Ed. by Starre, H. v. d.), pp. 397-400. London: IRL Press.
Kissler, J. & Bäuml, K.-H. 2000. Effects of the beholder's age on the perception of facial attractiveness. *Acta Psychologica*, 104(2), 145-166.
Kitamura, C., Thanavishuth, C., Burnhama, D. & Luksaneeyanawin, S. 2002. Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behavior & Development*, 24, 372-392.
Klein, F., Lemaire, V., Sandi, C., Vitiello, S., Van der Logt, J., Laurent, P. E., Neveu, P., Le Moal, M. & Mormede, P. 1992. Prolonged increase of corticosterone secretion by chronic social stress does not necessarily impair immune functions. *Life Sciences*, 50(10), 723-731.
Klimkeit, E. I., Mattingley, J. B., Sheppard, D. M., Lee, P. & Bradshaw, J. L. 2003. Perceptual asymmetries in normal children and children with attention deficit/hyperactivity disorder. *Brain and Cognition*, 52(2), 205-215.
Knecht, M., Kuhnau, D., Huttenbrink, K.-B., Martin, W. & Hummel, T. 2001. Frequency and localization of the putative

vomeronasal organ in humans in relation to age and gender. *Laryngoscope*, 111(3), 448-452.

Kniffin, K. M. & Wilson, D. S. 2004. The effect of nonphysical traits on the perception of physical attractiveness: Three naturalistic studies. *Evolution and Human Behavior*, 25(2), 88.

Koehler, N., Rhodes, G. & Simmons, L. W. 2002. Are human female preferences for symmetrical male faces enhanced when conception is likely? *Animal Behaviour*, 64(2), 233.

Kohl, J. V., Atzmueller, M., Fink, B. & Grammer, K. 2001. Human pheromones: Integrating neuroendocrinology and ethology. *Neuroendocrinology Letters*, 22(5), 309-321.

Komlos, L., Zamir, R., Joshua, H. & Halbrecht, I. 1977. Common HLA antigens in couples with repeated abortions. *Clinical Immunology and Immunopathology*, 7, 330-335.

Kostyu, D. D., Dawson, D. V., Elias, S. & Ober, C. 1993. Deficit of HLA homozygotes in a Caucasian isolate. *Human Immunology*, 37(3), 135-142.

Kouro-Mehr, H., Pintchovski, S., Melnyk, J., Chen, Y. J., Friedman, C., Trask, B. & Shizuya, H. 2001. Identification of non-functional human VNO receptor genes provides evidence for vestigiality of the human VNO *Chemical Senses*, 26, 1167-1174.

Kovacs, G., Gulyas, B. C., Savic, I., Perrett, D. I., Cornwell, R. E., Little, A. C., Jones, B. C., Burt, D. M., Gal, V. & Vidnyanszky, Z. 2004. Smelling human sex hormone-like compounds affects face gender judgment of men. *Neuroreport*, 15(8), 1275-1277.

Krantz, J. H., Ballard, J. & Scher, J. 1997. Comparing the results of laboratory and World-Wide Web samples on the determinants of female attractiveness. *Behavior Research Methods, Instruments, & Computers*, 29(2), 264-269.

Krebs, J. R. & Dawkins, R. 1984. Animal signals: Mind-reading and manipulation. In: *Behavioural ecology: An evolutionary approach* (Ed. by Krebs, J. R. & Davies, N. B.), pp. 380 - 402. Oxford: Blackwell.

Krug, R., Mölle, M., Fehm, H. L. & Born, J. 1999. Variations across the menstrual cycle in EEG activity during thinking and mental relaxation. *Journal of Psychophysiology*, 13(3), 163-172.

Kuhn, T. S. 1962. The Structure of Scientific Revolutions. Chicago: University of Chicago Press.

Kuukasjarvi, S., Eriksson, C. J. P., Koskela, E., Mappes, T., Nissinen, K. & Rantala, M. J. 2004. Attractiveness of women's body odors over the menstrual cycle: the role of oral contraceptives and receiver sex. *Behavioral Ecology*, 15(4), 579-584.

Laeng, B. & Falkenberg, L. 2007. Women's pupillary responses to sexually significant others during the hormonal cycle. *Hormones and Behavior*, 52(4), 520-530.

Laland, K. N. & Brown, G. 2002. Sense and Nonsense: Evolutionary Perspectives on Human Behaviour. Oxford: Oxford University Press.

Lander, K. 2008. Relating visual and vocal attractiveness for moving and static faces. *Animal Behaviour*, 75(3), 817-822.

Langlois, J. H. & Roggman, L. A. 1990. Attractive faces are only average. *Psychological Science*, 1, 115-121.

Langlois, J. H., Ritter, J. M., Roggman, L. A. & Vaughn, L. S. 1991. Facial diversity and infant preferences for attractive faces. *Developmental Psychology*, 27(1), 79-84.

Langlois, J. H., Roggman, L. A. & Musselman, L. 1994. What is average and what is not average about attractive faces? *Psychological Science*, 5, 214-219.

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M. & Smoot, M. 2000. Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, 126(3), 390-423.

Law Smith, M. J., Perrett, D. I., Jones, B. C., Cornwell, R. E., Moore, F. R., Feinberg, D. R., Boothroyd, L. G., Durrani, S. J., Stirrat, M. R., Whiten, S., Pitman, R. M. & Hillier, S. G. 2007. Facial appearance is a cue to oestrogen levels in women. *Proceedings of the Royal Society B: Biological Sciences*, 273(1583), 135-140.

Lee, P. A. & Migeon, C. J. 1975. Puberty in boys: correlation of plasma levels of gonadotropins (LH, FSH), androgens (testosterone, androstenedione, dehydroepiandrosterone and its sulfate), estrogens (estrone and estradiol) and progestins (progesterone and 17-hydroxyprogesterone). *Journal of Clinical Endocrinology and Metabolism*, 41, 556-562.

Lee, S., Potamianos, A. & Narayanan, S. 1999. Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, 105(3), 1455-1468.

Lein, A. 1979. The cycling female: her menstrual rhythm. San Francisco: Freeman.

Lens, L., Van Dongen, S., Kark, S. & Matthysen, E. 2002. Fluctuating asymmetry as an indicator of fitness: can we bridge the gap between studies? *Biological Reviews*, 77, 27-38.

Leung, B. & Forbes, M. R. 1996. Fluctuating asymmetry in relation to stress and fitness: Effects of trait type as revealed by meta-analysis. *Ecoscience*, 3(4), 400-413.

Li, N. P. & Kenrick, D. T. 2006. Sex similarities and differences in preferences for short-term mates: What, whether, and why. *Journal of Personality and Social Psychology*, 90(3), 468-489.

Li, W., Moallem, I., Paller, K. A. & Gottfried, J. A. 2007. Subliminal smells can guide social preferences. *Psychological Science*, 18(12), 1044-1049.

Liberles, S. D. & Buck, L. B. 2006. A second class of chemosensory receptors in the olfactory epithelium. *Nature*, 442, 645-650.

Lipson, S. F. & Ellison, P. T. 1996. Comparison of salivary steroid profiles in naturally occurring conception and nonconception cycles. *Human Reproduction*, 11, 2090-2096.

Little, A. C., Burt, D. M., Penton-Voak, I. S. & Perrett, D. I. 2001. Self-perceived attractiveness influences human female preferences for sexual dimorphism and symmetry in male faces. *Proceedings of the Royal Society - Biological Sciences (Series B)*, 268(1462), 39-44.

Little, A. C. & Hancock, P. J. 2002. The role of masculinity and distinctiveness on the perception of attractiveness in human male faces. *British Journal of Psychology*, 93(4), 451-464.

Little, A. C., Jones, B. C., Penton-Voak, I. S., Burt, D. M. & Perrett, D. I. 2002. Partnership status and the temporal context of relationships influence human female preferences for sexual dimorphism in male face shape. *Proceedings of the Royal Society - Biological Sciences (Series B)*, 269(1496), 1095-1100.

Little, A. C. & Jones, B. C. 2003. Evidence against perceptual bias views for symmetry preferences in human faces. *Proceedings Of The Royal Society Of London Series B-Biological Sciences*, 270(1526), 1759-1763.

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.

Little, A. C., DeBruine, L. M. & Jones, B. C. 2005. Sex-contingent face after-effects suggest distinct neural populations code male and female faces. *Proceedings Of The Royal Society Of London Series B-Biological Sciences*, 272(1578), 2283 - 2287.

Little, A. C. & Mannion, H. 2006. Viewing attractive or unattractive same-sex individuals changes self-rated attractiveness and face preferences in women. *Animal Behaviour*, 72(5), 981.

Little, A. C., Cohen, D., Jones, B. & Belsky, J. 2007a. Human preferences for facial masculinity change with relationship type and environmental harshness. *Behavioral Ecology and Sociobiology*, 61(6), 967-973.

Little, A. C., Jones, B. C. & Burriss, R. P. 2007b. Preferences for masculinity in male bodies change across the menstrual cycle. *Hormones and Behavior*, 51(5), 633.

Little, A. C., Jones, B. C., Waitt, C., Tiddeman, B. P., Feinberg, D. R., Perrett, D. I., Apicella, C. L. & Marlowe, F. W. 2008. Symmetry is related to sexual dimorphism in faces: Data across culture and species. *PLoS ONE*, 3(5), e2106.

Lopez, S. 1999. Parasitized female guppies do not prefer showy males. *Animal Behaviour*, 57, 1129-1134.

Lorenzen, L. A., Grieve, F. G. & Thomas, A. 2004. Brief report: Exposure to muscular male models decreases men's body satisfaction. *Sex Roles*, 51(11), 743-748.

Lundström, J. N., Goncalves, M., Esteves, F. & Olsson, M. J. 2003a. Psychological effects of subthreshold exposure to the putative human pheromone 4,16-androstadien-3-one. *Hormones and Behavior*, 44(5), 395-401.

Lundström, J. N., Hummel, T. & Olsson, M. J. 2003b. Individual differences in sensitivity to the odor of 4,16-androstadien-3-one. *Chemical Senses*, 28(7), 643-650.

Lundström, J. N. & Olsson, M. J. 2005. Subthreshold amounts of social odorant affect mood, but not behavior, in heterosexual women when tested by a male, but not a female, experimenter. *Biological Psychology*, 70(3), 197-204.

Lundström, J. N., Olsson, M. J., Schaal, B. & Hummel, T. 2006. A putative social chemosignal elicits faster cortical responses than perceptually similar odorants. *Neuroimage*, 30(4), 1340.

Maes, H. H., Neale, M. C., Kendler, K. S., Hewitt, J. K., Silberg, J. L., Foley, D. L., Meyer, J. M., Rutter, M., Simonoff, E., Pickles, A. & Eaves, L. J. 1998. Assortative mating for major psychiatric diagnoses in two population-based samples. *Psychological Medicine*, 28(6), 1389 - 1401.

Malina, R. M., Selby, H. A., Buschang, P. H., Aronson, W. L. & Little, B. B. 1983. Assortative mating for phenotypic characteristics in a Zapotec community in Oaxaca, Mexico. *Journal of Biosocial Science*, 15, 273-280.

Mallet, A. I., Holland, K. T., Rennie, P. J., Watkins, W. J. & Gower, D. B. 1991. Applications of gas chromatography--mass spectrometry in the study of androgen and odorous 16-androstene metabolism by human axillary bacteria. *Journal of Chromatography: Biomedical Applications*, 562(1-2), 647-658.

Manning, J. T., Scutt, D., Whitehouse, G. H., Leinster, S. J. & Walton, J. M. 1996. Asymmetry and the menstrual cycle in women. *Ethology and Sociobiology*, 17(2), 129-143.

Marshall, W. A. & Tanner, J. M. 1969. Variations in the pattern of pubertal changes in girls. *Archives of Disease in Childhood*, 44, 291-303.

Marshall, W. A. & Tanner, J. M. 1970. Variations in the pattern of pubertal changes in boys. *Archives of Disease in Childhood*, 45(239), 13-23.

Mathes, E. & Kozak, G. 2008. The exchange of physical attractiveness for resource potential and commitment. *Journal of Evolutionary Psychology*, 6(1), 43-56.

Mathews, C. A. & Reus, V. I. 2001. Assortative mating in the affective disorders: a systematic review and meta-analysis. *Comprehensive Psychiatry*, 42(4), 257 - 262.

Maynard Smith, J. 1979. Game theory and the evolution of behaviour. *Proceedings Of The Royal Society Of London Series B-Biological Sciences*, 205, 475-488.

Maynard Smith, J. 1982. Evolution and the theory of games. Cambridge: Cambridge University Press.

Maynard Smith, J. 1985. Mini review: Sexual selection, handicaps and true fitness. *Journal of Theoretical Biology*, 115, 1-8.

Mays, H. L. J. & Hill, G. E. 2004. Choosing mates: good genes versus genes that are a good fit. *Trends in Ecology and Evolution*, 19(10), 554-559.

Mazur, A. & Booth, A. 1998. Testosterone and dominance in men. *Behavioral and Brain Sciences*, 21(3), 353-397.

Mazzi, D., Künzler, R. & Bakker, T. C. M. 2003. Female preference for symmetry in computer-animated three-spined sticklebacks, Gasterosteus aculeatus. *Behavioral Ecology and Sociobiology*, 54(2), 156-161.

McClintock, M. K. 1971. Menstrual synchrony and suppression. *Nature*, 229, 244-245.

McClintock, M. K. & Herdt, G. 1996. Rethinking puberty: The development of sexual attraction. *Current Directions In Psychological Science*, 5(6), 178-183.

McClintock, M. K. 1998. Whither menstrual synchrony? *Annual Review of Sex Research*, 9, 77-95.

McClintock, M. K. 2000. Human pheromones: Primers, releasers, signallers, or modulators? In: *Reproduction in Context* (Ed. by Wallen, K. & Schneider, E.), pp. 335-420. Cambridge, MA: MIT Press.

McClintock, M. K., Jacob, S., Zelano, B. & Hayreh, D. J. 2001. Pheromones and vasanas: the functions of social chemosignals. *Nebraska Symposium on Motivation*, 47, 75-112.

McClure, E. B. 2000. A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin*, 126(3), 424-453.

McCollough, P. A., Owen, J. W. & Pollak, E. I. 1981. Does androstenol affect emotion? *Ethology and Sociobiology*, 2(2), 85-88.

McFarland, D. 1993. Animal Behavior: Longman Scientific and Technical.

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.

Mealey, L., Bridgstock, R. & Townsend, G. C. 1999. Symmetry and perceived facial attractiveness: a monozygotic co-twin comparison. *Journal of Personality and Social Psychology*, 76, 157-165.

Meissner, C. A. & Brigham, J. C. 2001. Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy and Law*, 7(1), 3-35.

Meredith, M. 2001. Human vomeronasal organ function: A critical review of best and worst cases *Chemical Senses*, 26, 433-445.

Meredydd, H., Sarah, H., Jeremy, H. & Ieuan, H. 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 and Otology*, 112(5), 451-454.

Merikangas, K. R. 1982. Assortative mating for psychiatric disorders and psychological traits. *Archives of General Psychiatry*, 39(10), 1173-1180.

Miller, G. F., Tybur, J. M. & Jordan, B. D. 2007. Ovulatory cycle effects on tip earnings by lap dancers: economic evidence for human estrus? *Evolution and Human Behavior*, 28(6), 375-381.

Miller, P. B. & Soules, M. R. 1996. The usefulness of a urinary LH kit for ovulation prediction during menstrual cycles of normal women. *Obstetrics and Gynecology*, 87(1), 13-17.

Miyake, K. & Zuckerman, M. 1993. Beyond personality impressions: Effects of physical and vocal attractiveness on false consensus, social comparison, affiliation, and assumed and perceived similarity. *Journal of Personality*, 61(3), 411-437.

Møller, A. P. & Pomiankowski, A. 1993. Why have birds got multiple sexual ornaments? *Behavioral Ecology and Sociobiology*, 32(3), 167-176.

Møller, A. P. 1997. Developmental stability and fitness: a review. *The American Naturalist*, 149(5), 916-932.

Money, J. & Ehrhardt, A. 1972. Man, Woman, Boy, Girl. Baltimore: John Hopkins University Press.

Monti-Bloch, L. & Grosser, B. I. 1991. Effect of putative pheromones on the electrical activity of the human vomeronasal organ and olfactory epithelium. *Journal of Steroid Biochemistry and Molecular Biology*, 39(4B), 573-582.

Moran, D. T., Jafek, B. W. & Rowley, J. C. I. 1991. The vomeronasal (Jacob's) organ in man: ultrastructure and frequency of occurrence. *Journal of Steroid Biochemistry and Molecular Biology*, 39, 545-552.

Moravec, M. L., Striedter, G. F. & Burley, N. T. 2006. Assortative pairing based on contact call similarity in budgerigars, Melopsittacus undulatus. *Ethology*, 112(11), 1108-1116.

Morrell, M. J., Giudice, L., Flynn, K. L., Seale, C. G., Paulson, A. J., Doñe, S., Flaster, E., Ferin, M. & Sauer, M. V. 2002. Predictors of ovulatory failure in women with epilepsy. *Annals of Neurology*, 52(6), 704-711.

Morris, N. M. & Udry, J. R. 1980. Validation of a self-administered instrument to assess stage of adolescent development. *Journal of Youth and Adolescence*, 9(3), 271-280.

Morrison, E. R., Gralewski, L., Campbell, N. & Penton-Voak, I. S. 2007. Facial movement varies by sex and is related to attractiveness. *Evolution and Human Behavior*, 28(3), 186-192.

Morton, J. & Johnson, M. H. 1991. CONSPEC and CONLERN: A two-process theory of infant face recognition. *Psychological Review*, 98, 164-181.

Mueser, K. T., Grau, B. W., Sussman, S. & Rosen, A. J. 1984. You're only as pretty as you feel: Facial expression as a determinant of physical attractiveness. *Journal of Personality and Social Psychology*, 46(2), 469-478.

Neff, B. D. & Pitcher, T. E. 2005. Genetic quality and sexual selection: an integrated framework for good genes and compatible genes. *Molecular Ecology*, 14(1), 19-38.

Nettle, D. 2006. The evolution of personality variation in humans and other animals. *American Psychologist*, 61(6), 622–631.

New, M. I., Levine, L. S. & Pang, S. 1981. Adrenal androgens and growth. In: *The Biology of Normal Human Growth: Transactions of the First Karolinska Institute Nobel Conference* (Ed. by Ritzen, M.), pp. 285-295. New York: Raven Press.

Nicholls, M. E. R., Bradshaw, J. L. & Mattingley, J. B. 1999. Free-viewing perceptual asymmetries for the judgement of brightness, numerosity and size. *Neuropsychologia*, 37(3), 307-314.

Nixon, A., Mallet, A. I. & Gower, D. B. 1988. Simultaneous quantification of five odorous steroids (16-androstenes) in the axillary hair of men *Journal of Steroid Biochemistry*, 29(5), 505-510.

Nordlander, C., Hammarström, L. & Lindblom, B. 1983. No role of HLA in mate selection. *Immunogenetics*, 18, 429-431.

Nowak, M. A., Tarczy-Hornoch, K. & Austyn, J. M. 1992. The optimal number of major histocompatibility complex molecules in an individual. *Proceedings of the National Academy of Sciences of the United States of America*, 89, 10896-10899.

Ober, C. 1995. HLA and reproduction: Lessons from studies in the Hutterites. *Placenta*, 16, 569-577.

Ober, C., Weitkamp, L. R., Cox, N., Dytch, H., Kostyu, D. & Elias, S. 1997. HLA and mate choice in humans. *American Journal of Human Genetics*, 61(3), 497-504.

Ober, C., Hyslop, T., Elias, S., Weitkamp, L. R. & Hauck, W. W. 1998. Human leukocyte antigen matching and fetal loss: results of a 10 year prospective study. *Human Reproduction*, 13(1), 33-38.

Oetting, S., Prove, E. & Bischof, H.-J. 1995. Sexual imprinting as a two-stage process: mechanisms of information storage and stabilization. *Animal Behaviour*, 50(2), 393.

Oguchi, T. & Kikuchi, H. 1997. Voice and interpersonal attraction. *Japanese Psychological Research*, 39(1), 56-61.

Oinonen, K. A. & Mazmanian, D. 2007. Facial symmetry detection ability changes across the menstrual cycle. *Biological Psychology*, 75(2), 136.

Olejnik, S. F. & Algina, J. 1984. Parametric ANCOVA and the rank transform ANCOVA when the data are conditionally non-normal and heteroscedastic. *Journal of Educational Statistics*, 9(2), 129.

Orr, D. P. & Ingersoll, G. 1995. The contribution of level of cognitive complexity and pubertal timing to behavioral risk in young adolescents. *Pediatrics*, 95(4), 528-533.

Parker, L. L. C., Penton-Voak, I. S., Attwood, A. S. & Munafo, M. R. 2008. Effects of acute alcohol consumption on ratings of attractiveness of facial stimuli: evidence of long-term encoding. *Alcohol Alcohol.*, agn065.

Parker, L. N., Sack, J., Fisher, D. A. & Odell, W. D. 1978. Adrenarche: Prolactin, gonadotropins, adrenal androgens, and cortisol. *Journal of Clinical Endocrinology and Metabolism*, 46(3), 396-401.

Parker, L. N. 1991. Adrenarche. *Endocrinology and Metabolism Clinics of North America*, 20(1), 71-83.

Pause, B. M. 2004. Are androgen steroids acting as pheromones in humans? *Physiology and Behavior*, 83(1 SPEC. ISS.), 21-29.

Pawlowski, B. & Dunbar, R. I. M. 1999. Impact of market value on human mate choice decisions. *Proceedings Of The Royal Society Of London Series B-Biological Sciences*, 266(1416), 281-285.

Pawlowski, B. & Jasienska, G. 2008. Women's body morphology and preferences for sexual partners' characteristics. *Evolution and Human Behavior*, 29(1), 19-25.

Peckham, P. D. 1968. An investigation of the effects of non-homogeneity of regression slopes upon the F-test of analysis of covariance. Report No. 16. Boulder, Colorado: Laboratory of Educational Research, University of Colorado.

Penn, D. J. & Potts, W. K. 1998. How do major histocompatibility complex genes influence odor and mating preferences? *Advances in Immunology*, 69, 411-435.

Penn, D. J. & Potts, W. K. 1999. The evolution of mating preferences and major histocompatibility genes. *American Naturalist*, 153, 145-164.

Penn, D. J. 2002. The scent of genetic compatibility: Sexual selection and the major histocompatibility complex.. *Ethology*, 108(1), 1-21.

Penton-Voak, I. S., Perrett, D. I., Castles, D. L., Kobayashi, T., Burt, D. M., Murray, L. K. & Minamisawa, R. 1999a. Menstrual cycle alters face preference. *Nature*, 399(6738), 741-742.

Penton-Voak, I. S., Perrett, D. I. & Peirce, J. W. 1999b. Computer graphic studies of the role of facial similarity in judgements of attractiveness. *Current Psychology: Developmental Learning Personality*, 18(1), 104-117.

Penton-Voak, I. S. & Perrett, D. I. 2000. Female preference for male faces changes cyclically: Further evidence. *Evolution and Human Behavior*, 21(1), 39.

Penton-Voak, I. S., Jones, B. C., Little, A. C., Baker, S., Tiddeman, B., Burt, D. M. & Perrett, D. I. 2001. Symmetry, sexual dimorphism in facial proportions and male facial attractiveness. *Proceedings of the Royal Society - Biological Sciences (Series B)*, 268(1476), 1617-1625.

Penton-Voak, I. S., Little, A. C., Jones, B. C., Burt, D. M., Tiddeman, B. P. & Perrett, D. I. 2003. Female condition influences preferences for sexual dimorphism in faces of male humans (*Homo sapiens*). *Journal of Comparative Psychology*, 117(3), 264-271.

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., Jacobson, A. & Trivers, R. L. 2004. Populational differences in attractiveness judgements of male and female faces: Comparing British and Jamaican samples. *Evolution and Human Behavior*, 25(6), 355.

Penton-Voak, I. S. & Chang, H. 2008. Attractiveness judgements of individuals vary across emotional expression and movement conditions. *Journal of Evolutionary Psychology*, 6(2), 89-100.

Perrett, D. I., May, K. A. & Yoshikawa, S. 1994. Facial shape and judgments of female attractiveness. *Nature*, 368, 239-242.

Perrett, D. I., Lee, K. J., Penton-Voak, I. S., Rowland, D., Yoshikawa, S., Burt, D. M., Henzi, S. P., Castles, D. L. & Akamatsu, S. 1998. Effects of sexual dimorphism on facial attractiveness. *Nature*, 394(6696), 884-887.

Perrett, D. I., Burt, D. M., Penton-Voak, I. S., Lee, K. J., Rowland, D. A. & Edwards, R. 1999. Symmetry and human facial attractiveness. *Evolution and Human Behavior*, 20, 295-307.

Perrett, D. I., Penton-Voak, I. S., Little, A. C., Tiddeman, B. P., Burt, D. M., Schmidt, N., Oxley, R., Kinloch, N. & Barrett, L. 2002. Facial attractiveness judgements reflect learning of parental age characteristics. *Proceedings Of The Royal Society Of London Series B-Biological Sciences*, 269(1494), 873-880.

Peters, M., Rhodes, G. & Simmons, L. W. 2007. Contributions of the face and body to overall attractiveness. *Animal Behaviour*, 73(6), 937-942.

Peters, M., Simmons, L. W. & Rhodes, G. 2008. Testosterone is associated with mating success but not attractiveness or masculinity in human males. *Animal Behaviour*, In Press, Corrected Proof.

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, 117-133.

Petrie, M., Krupa, A. & Burke, T. 1999. Peacocks lek with relatives even in the absence of social and environmental cues. *Nature*, 401(6749), 155.

Pipitone, R. N. & Gallup, G. G. J. 2008. Women's voice attractiveness varies across the menstrual cycle. *Evolution and Human Behavior*, 29(4), 268-274.

Plant, T. M. 1986. Gonadal regulation of hypothalamic gonadotropin-releasing hormone release in primates. *Endocrine Reviews*, 7, 75-88.

Plant, T. M. 2002. Neurophysiology of puberty. *Journal of Adolescent Health*, 31, 185–191.

Platek, S. M., Burch, R. L., Panyavin, I. S., Wasserman, B. H. & Gallup, J. G. G. 2002. Reactions to children's faces: Resemblance affects males more than females. *Evolution and Human Behavior*, 23(3), 159.

Pollard, J., Shepherd, J. & Shepherd, J. 1999. Average faces are average faces. *Current Psychology*, 18(1), 98-103.

Poran, N. S. 1994. Cycle attractivity of human female odors. *Advances in the Biosciences*, 93, 555-560.

Potts, W. K. & Wakeland, E. K. 1993. Evolution of MHC genetic diversity - a tale of incest, pestilence and sexual preference. *Trends in Genetics*, 9(12), 408-412.

Potts, W. K., Manning, C. J. & Wakeland, E. K. 1994. The role of infectious disease, inbreeding, and mating preferences in maintaining MHC diversity: an experimental test. *Philosophical Transactions of the Royal Society of London, Series B*, 346, 369-378.

Provost, M. P., Quinsey, V. L. & Troje, N. F. 2008. Differences in gait across the menstrual cycle and their attractiveness to men. *Archives of Sexual Behavior.*

Puts, D. A. 2005. Mating context and menstrual phase affect women's preferences for male voice pitch. *Evolution and Human Behavior*, 26(5), 388-397.

Puts, D. A. 2006. Cyclic variation in women's preferences for masculine traits: Potential hormonal causes. *Human Nature*, 17(1), 114-127.

Quinn, P. C., Kelly, D. J., Lee, K., Pascalis, O. & Slater, A. M. 2008. Preference for attractive faces in human infants extends beyond conspecifics. *Developmental Science*, 11(1), 76-83.

Qvarnström, A. 2001. Context-dependent genetic benefits from mate choice. *Trends In Ecology & Evolution*, 16(1), 5-7.

Raines, R. S., Hechtman, S. B. & Rosenthal, R. 1990a. Physical attractiveness of face and voice: Effects of positivity, dominance, and sex. *Journal of Applied Social Psychology*, 20(19), 1558-1578.

Raines, R. S., Hechtman, S. B. & Rosenthal, R. 1990b. Nonverbal behavior and gender as determinants of physical attractiveness. *Journal of Nonverbal Behavior*, 14(4), 253-267.

Rantala, M. J., Vainikka, A. & Kortet, R. 2003. The role of juvenile hormone in immune function and pheromone production trade-offs: a test of the immunocompetence handicap principle. *Proceedings Of The Royal Society Of London Series B-Biological Sciences*, 270(1530), 2257-2261.

Read, A. F. & Harvey, P. H. 1991. Genetic relatedness and the evolution of animal mating patterns. In: *Human Mating Patterns* (Ed. by Mascie-Taylor, C. G. & Boyce, A. J.), pp. 115-131. Oxford: Oxford University Press.

Reiter, E. O. & Rosenfeld, R. G. 2003. Normal and aberrant growth. In: *Williams textbook of endocrinology* (Ed. by Larsen, P. R., Kronenberg, H. M., Melmed, S. & Polonsky, K. S.), pp. 1003-1114. New York: Elsevier.

Renaud, R. L., Macler, J., Dervain, I., Ehret, M. C., Aron, C., Plas-Roser, S., Spira, A. & Pollack, H. 1980. Echographic study of follicular maturation and ovulation during the normal menstrual cycle. *Fertility and Sterility*, 33(3), 272-276.

Restrepo, D., Arellano, J., Oliva, A. M., Schaefer, M. L. & Lin, W. 2004. Emerging views on the distinct but related roles of the main and accessory olfactory systems in responsiveness to chemosensory signals in mice. *Hormones and Behavior*, 46, 247-256.

Rhodes, G., Proffitt, F., Grady, J. & Sumich, A. 1998. Facial symmetry and the perception of beauty. *Psychonomic Bulletin and Review*, 5, 659-669.

Rhodes, G., Sumich, A. & Byatt, G. 1999. Are average facial configurations attractive only because of their symmetry? *Psychological Science*, 10(1), 52-58.

Rhodes, G., Hickford, C. & Jeffery, L. 2000. Sex-typicality and attractiveness: Are supermale and superfemale faces super-attractive? *British Journal of Psychology*, 91(1), 125.

Rhodes, G., Halberstadt, J. & Brajkovick, G. 2001. Generalization of mere exposure effects to averaged composite faces. *Social Cognition*, 19(1), 57-70.

Rhodes, G., Geddes, K., Jeffery, L., Dziurawiecz, S. & Clark, A. 2002. Are average and symmetric faces attractive to infants? Discrimination and looking preferences. *Perception*, 31, 315-321.

Rhodes, G., Chan, J., Zebrowitz, L. A. & Simmons, L. W. 2003a. Does sexual dimorphism in human faces signal health? *Biology Letters*, 270(Suppl.), S93–S95.

Rhodes, G., Jeffery, L., Watson, T. L., Clifford, C. W. G. & Nakayama, K. 2003b. Fitting the mind to the world: face adaptation and attractiveness aftereffects. *Psychological Science*, 14(6), 558-566.

Rhodes, G. 2006. The evolutionary psychology of facial beauty. *Annual Review of Psychology*, 57, 199-226.

Rhodes, G., Yoshikawa, S., Palermo, R., Simmons, L. W., Peters, M., Lee, K., Halberstadt, J. & Crawford, J. R. 2007. Perceived health contributes to the attractiveness of facial symmetry, averageness, and sexual dimorphism. *Perception*, 36(8), 1244-1252.

Riding, D., Lonsdale, D. & Brown, B. 2006. The effects of average fundamental frequency and variance of fundamental frequency on male vocal attractiveness to women *Journal of Nonverbal Behavior*, 30(2), 55-61.

Riggio, R. E., Widaman, K. F., Tucker, J. S. & Salinas, C. 1991. Beauty is more than skin deep: Components of attractiveness. *Basic and Applied Social Psychology*, 12(4), 423-439.

Rikowski, A. & Grammer, K. 1999. Human body odour, symmetry and attractiveness. *Proceedings of the Royal Society - Biological Sciences (Series B)*, 266(1422), 869-874.

Roberts, S. C. & Gosling, L. M. 2003. Genetic similarity and quality interact in mate choice decisions by female mice. *Nature Genetics*, 35(1), 103-106.

Roberts, S. C., Havlicek, J., Flegr, J., Hruskova, M., Little, A. C., Jones, B. C., Perrett, D. I. & Petrie, M. 2004. Female facial attractiveness increases during the fertile phase of the menstrual cycle. *Proceedings of the Royal Society B: Biological Sciences*, 271(Biology Letters Supplement 5), S270-272.

Roberts, S. C., Gosling, L. M., Spector, T. D., Miller, P., Penn, D. J. & Petrie, M. 2005a. Body odor similarity in noncohabiting twins. *Chemical Senses*, 30, 1-6.

Roberts, S. C., Little, A. C., Gosling, L. M., Jones, B. C., Perrett, D. I., Carter, V. & Petrie, M. 2005b. MHC-assortative facial preferences in humans. *Biology Letters*, 1(4), 400-403.

Roberts, S. C., Little, A. C., Gosling, L. M., Perrett, D. I., Carter, V., Jones, B. C., Penton-Voak, I. & Petrie, M. 2005c. MHC-heterozygosity and human facial attractiveness. *Evolution and Human Behavior*, 26, 213-226.

Roberts, S. C. & Little, A. C. 2008. Good genes, complementary genes and human mate choice. *Genetica*, 132, 309-321.

Roberts, S. C., Saxton, T. K., Murray, A., Burriss, R., Rowland, H. & Little, A. C. in review. Static and dynamic facial images cue similar attractiveness judgements. *Ethology*.

Rodriguez, C., Patel, A. V., Calle, E. E., Jacob, E. J. & Thun, M. J. 2001. Estrogen replacement therapy and ovarian cancer mortality in a large prospective study of US women. *The Journal of the American Medical Association*, 285, 1460-1465.

Roney, J. R. & Simmons, Z. L. 2008. Women's estradiol predicts preference for facial cues of men's testosterone. *Hormones and Behavior*, 53(1), 14-19.

Rosenberg, L. T., Cooperman, D. & Payn, R. 1983. HLA and mate selection. *Immunogenetics*, 17(1), 89-93.

Rowe, C. 1999. Receiver psychology and the evolution of multicomponent signals. *Animal Behaviour*, 58(5), 921-931.

Rowe, R., Maughan, B., Worthman, C. M., Costello, E. J. & Angold, A. 2004. Testosterone, antisocial behavior, and social dominance in boys: pubertal development and biosocial interaction. *Biological Psychiatry*, 55(5), 546.

Rubenstein, A. J. 2005. Variation in perceived attractiveness - Differences between dynamic and static faces. *Psychological Science*, 16(10), 759-762.

Russell, M. J. 1976. Human olfactory communication. *Nature*, 260, 520-522.

Samuels, C. A., Butterworth, G., Roberts, T., Grauper, L. & Hole, G. 1994. Facial aesthetics: babies prefer attractiveness to symmetry. *Perception*, 23, 823-831.

Savic, I., Berglund, H., Gulyas, B. & Roland, P. 2001. Smelling of odorous sex hormone-like compounds causes sex-differentiated hypothalamic activations in humans. *Neuron*, 31(4), 661-668.

Savic, I., Berglund, H. & Lindstrom, P. 2005. Brain response to putative pheromones in homosexual men. *Proceedings of the National Academy of Sciences of the United States of America*, 102(20), 7356-7361.

Saxton, T. K., Caryl, P. G. & Roberts, S. C. 2006. Vocal and facial attractiveness judgments of children, adolescents and adults: the ontogeny of mate choice. *Ethology*, 112, 1179-1185.

Saxton, T. K., Little, A. C. & Roberts, S. C. 2007. Ecological validity in the study of human pheromones. In: *Chemical Signals in Vertebrates* (Ed. by Hurst, J. L., Beynon, R.J., Roberts, S.C., Wyatt, T.J.), pp. 111-120. New York: Springer.

Saxton, T. K., Lyndon, A., Little, A. C. & Roberts, S. C. 2008. Evidence that androstadienone, a putative human chemosignal, modulates women's attributions of men's attractiveness. *Hormones and Behavior*, 54, 597-601.

Saxton, T. K., Little, A. C., Rowland, H., Gao, T. & Roberts, S. C. in prep. Women's face-based mating preferences trade off absolute and relative quality.

Saxton, T. K., Burriss, R., Murray, A., Rowland, H. & Roberts, S. C. in review-a. Visual and vocal cues independently predict judgments of overall attractiveness. *Journal of Evolutionary Psychology*.

Saxton, T. K., DeBruine, L. M., Jones, B. C., Little, A. C. & Roberts, S. C. in review-b. Face and voice attractiveness judgments change during adolescence. *Evolution and Human Behavior*.

Saxton, T. K., Gao, T., Burriss, R., Rowland, H., Havlicek, J. & Roberts, C. R. in review-c. Changes in attractiveness across the menstrual cycle are detectable in multiple modalities. *Behavioral Ecology and Sociobiology*.

Scarbrough, P. S. & Johnston, V. S. 2005. Individual differences in women's facial preferences as a function of digit ratio and mental rotation ability. *Evolution and Human Behavior*, 26(6), 509.

Schachter, B., Weitkamp, L. R. & Johnson, W. E. 1984. Parental HLA compatibility, fetal wastage and neural tube defects: evidence for a T/t-like locus in humans. *American Journal of Human Genetics*, 36, 1082-1091.

Scheib, J. E., Gangestad, S. W. & Thornhill, R. 1999. Facial attractiveness, symmetry and cues of good genes. *Proceedings of the Royal Society B Biological Sciences*, 266(1431), 1913-1917.

Schmalt, H.-D. 2006. Waist-to-hip ratio and female physical attractiveness: The moderating role of power motivation and the mating context. *Personality And Individual Differences*, 41(3), 455-465.

Schmitt, D. P. 2005. Sociosexuality from Argentina to Zimbabwe: A 48-nation study of sex, culture, and strategies of human mating. *Behavioral and Brain Sciences*, 28, 247-311.

Schwarz, S. & Hassebrauck, M. 2008. Self-perceived and observed variations in women's attractiveness throughout the menstrual cycle--a diary study. *Evolution and Human Behavior*, 29(4), 282-288.

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

Scott, I., Swami, V., Josephson, S. C. & Penton-Voak, I. S. 2008. Context-dependent preferences for facial dimorphism in a rural Malaysian population. *Evolution and Human Behavior*, 29(4), 289-296.

Scutt, D. & Manning, J. T. 1996. Ovary and ovulation: Symmetry and ovulation in women. *Human Reproduction*, 11(11), 2477-2480.

Service, R. F. 1998. New role for estrogen in cancer? *Science*, 279, 1631-1633.

Shackelford, T. K. & Larsen, R. J. 1997. Facial asymmetry as an indicator of psychological, emotional and physiological distress. *Journal of Personality and Social Psychology*, 72(2), 456-466.

Shackelford, T. K. & Larsen, R. J. 1999. Facial attractiveness and physical health. *Evolution and Human Behavior*, 20(1), 71-76.

Sherwood, L. 1989. Human Physiology from Cells to Systems. St Paul, MN: West Publishing Company.

Shinohara, K., Morofushi, M., Funabashi, T., Mitsushima, D. & Kimura, F. 2000. Effects of 5-androst-16-en-3-ol on the pulsatile secretion of luteinizing hormone in human females *Chemical Senses*, 25(4), 456-467.

Signoret, J. P. & du Mesnil du Buisson, F. 1961. Etude du comportement de la truie en oestrus. *Proceedings of the 4th International Congress on Animal Reproduction*, 2, 171-175.

Silverman, E. M. & Zimmer, C. H. 1978. Effect of the menstrual cycle on voice quality. *Archives of Otolaryngology*, 104, 7-10.

Simmons, R. G., Blyth, D. A., Cleave, E. F. V. & Bush, D. M. 1979. Entry into early adolescence: The impact of school structure, puberty, and early dating on self-esteem. *American Sociological Review*, 44(6), 948.

Singh, D. & Bronstad, P. M. 2001. Female body odour is a potential cue to ovulation. *Proceedings of the Royal Society - Biological Sciences (Series B)*, 268(1469), 797-801.

Sizonenko, P., Paunier, L. & Carmignac, D. 1976. Hormonal changes during puberty. IV. Longitudinal study of adrenal androgen secretions. *Hormone Research*, 74(4-5), 288-302.

Slater, A., Von der Schulenburg, C., Brown, E., Badenoch, M., Butterworth, G., Parsons, S. & Samuels, C. 1998. Newborn infants prefer attractive faces. *Infant Behavior and Development*, 21(2), 345-354.

Spuhler, J. N. 1968. Assortative mating with respect to physical characteristics. *Eugenics Quarterly*, 15(2), 128-140.

Steiner, D. D. & Rain, J. S. 1989. Immediate and delayed primacy and recency effects in performance evaluation. *Journal of Applied Psychology*, 74(1), 136-142.

Stensaas, L. J., Lakver, R. M., Monti-Bloch, L., Grosser, B. I. & Berliner, D. L. 1991. Ultrastructure of the human vomeronasal organ. *Journal of Steroid Biochemistry and Molecular Biology*, 39, 553-560.

Stewart, D. R., Overstreet, J. W., Nakajima, S. T. & Lasley, B. L. 1993. Enhanced ovarian steroid secretion before implantation in early human pregnancy. *Journal of Clinical Endocrinology and Metabolism*, 76, 1470-1476.

Stylianou, M., Forchielli, E., Tummillo, M. & Dorfman, R. I. 1961. Metabolism in vitro of 4-C14-testosterone by a human liver homogenate. *Journal of Biological Chemistry*, 236(3), 692-694.

Subrahmaniam, K., Subrahmaniam, K. & Messeri, J. Y. 1975. On the robustness of some tests of significance in sampling from a compound normal population. *Journal of the American Statistical Association*, 70(350), 435.

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

Swami, V. & Tovée, M. J. 2006. Does hunger influence judgments of female physical attractiveness? *British Journal of Psychology*, 97(3), 353-363.

Swami, V. & Tovée, M. 2007. Differences in attractiveness preferences between observers in low- and high-resource environments in Thailand. *Journal of Evolutionary Psychology*, 5(1), 149-160.

Symons, D. 1995. The Evolution of Human Sexuality. New York: Oxford University Press.

Tanner, J. M. 1962. Growth at Adolescence: With a general consideration of the effects of hereditary and environmental factors upon growth and maturation from birth to maturity. Oxford: Blackwell Scientific Publications.

Tanner, J. M. 1978. Foetus into Man: Physical Growth from Conception to Maturity. London: Open Books.

Taylor, Whincup, Hindmarsh, Lampe, Odoki & Cook. 2001. Performance of a new pubertal self-assessment questionnaire: a preliminary study. *Paediatric and Perinatal Epidemiology*, 15(1), 88-94.

Thamdrup, E. 1961. Precocious Sexual Development, A Clinical Study of 100 Children. Kopenhagen: Munksgaard.

Thiessen, D. & Gregg, B. 1980. Human assortative mating and genetic equilibrium: an evolutionary perspective. *Ethology and Sociobiology*, 1, 111-140.

Thiessen, D. 1999. Social influences on assortative mating. In: *The Descent of Mind: Psychological perspectives on hominid evolution* (Ed. by Corballis, M. C. & Lea, S. G.), pp. 311-323. Oxford: Oxford University Press.

Thorne, F., Neave, N., Scholey, A., Moss, M. & Fink, B. 2002. Effects of putative male pheromones on female ratings of male attractiveness: Influence of oral contraceptives and the menstrual cycle. *Neuroendocrinology Letters*, 23(4), 291-297.

Thornhill, R. & Gangestad, S. W. 1993. Human facial beauty: averageness, symmetry and parasite resistance. *Human Nature*, 4, 237-263.

Thornhill, R. & Gangestad, S. W. 1999a. Facial attractiveness. *Trends in Cognitive Sciences*, 3(12), 452.

Thornhill, R. & Gangestad, S. W. 1999b. The scent of symmetry: A human sex pheromone that signals fitness? *Evolution and Human Behavior*, 20, 175-201.

Thornhill, R. & Grammer, K. 1999. The body and face of woman: one ornament that signals quality? *Evolution and Human Behavior*, 20, 105-120.

Thornhill, R., Gangestad, S. W., Miller, R., Scheyd, G., McCollough, J. K. & Franklin, M. 2003. Major histocompatibility complex genes, symmetry, and body scent attractiveness in men and women. *Behavioral Ecology*, 14(5), 668-678.

Tiddeman, B., Burt, D. M. & Perrett, D. 2001. Computer graphics in facial perception research. *IEEE Computer Graphics and Applications*, 21(5), 42-50.

Timiras, P. 1972. Developmental Physiology and Aging. New York: Macmillan.

Tinbergen, N. 1963. On aims and methods in ethology. *Zeitschrift fur Tierpsychologie*, 20, 410-433.

Tirindelli, R., Mucignat-Caretta, C. & Ryba, N. J. 1998. Molecular aspects of pheromonal communication via the vomeronasal organ of mammals. *Trends in Neurosciences*, 21, 482-486.

Titze, I. R. 1994. Principles of Voice Production. New Jersey: Prentice Hall.

Tosi, O., Postan, D. & Bianculli, C. 1976. Longitudinal study of children's voices at puberty. In: *XVIth International Congress of Logopedics and Phoniatrics* (Ed. by Loebell, E.), pp. 486-490. Interlaken 1974: S. Karger.

Trainor, L. J. & Zacharias, C. A. 1998. Infants prefer higher-pitched singing. *Infant Behavior and Development*, 21(4), 799-806.

Tremblay, R. E., Schaal, B., Boulerice, B., Arseneault, L., Soussignan, R. G., Paquette, D. & Laurent, D. 1998. Testosterone, physical aggression, dominance, and physical development in early adolescence. *International Journal of Behavioral Development*, 22(4), 753-777.

Trivers, R. L. 1972. Parental investment and sexual selection. In: *Sexual Selection and the Descent of Man* (Ed. by Campbell, B.), pp. 136-179. London: Heinemann.

Trivers, R. L., Manning, J. T., Thornhill, R., Singh, D. & McGuire, M. 1999. Jamaican Symmetry Project: long-term study of fluctuating asymmetry in rural Jamaican children. *Human Biology*, 71(3), 417-430.

Trotier, D., Eloit, C., Wassef, M., Talmain, G., Bensimon, J. L., Doving, K. B. & Ferrand, J. 2000. The vomeronasal cavity in adult humans. *Chemical Senses*, 25(4), 369-380.

Udry, J. R., Billy, J. O. G., Morris, N. M., Gross, T. R. & Raj, M. H. 1985. Serum androgenic hormones motivate sexual behavior in boys. *Fertility and Sterility*, 43(1), 90-94.

Udry, J. R. & Billy, J. O. G. 1987. Initiation of coitus in early adolescence. *American Sociological Review*, 52(6), 841-855.

Udry, J. R. 1988. Biological predispositions and social control in adolescent sexual behavior. *American Sociological Review*, 53(5), 709-722.

Udry, J. R. 1990. Hormonal and social determinants of adolescent sexual initiation. In: *Adolescence and Puberty* (Ed. by Bancroft, J.), pp. 70 - 87. Oxford: Oxford University Press.

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 Anders, S. M. & Watson, N. V. 2007. Testosterone levels in women and men who are single, in long-distance relationships, or same-city relationships. *Hormones and Behavior*, 51(2), 286-291.

Van Valen, L. 1973. A new evolutionary law. *Evolutionary Theory*, 1, 1-30.

Vandenbergh, J. G. 1983a. Pheromonal regulation of puberty. In: *Pheromones and Reproduction in Mammals* (Ed. by Vandenbergh, J. G.), pp. 95-112. New York: Academic Press.

Vandenbergh, J. G. 1983b. *Pheromones and Reproduction in Mammals*. New York: Academic Press.

Vermeersch, H., T'Sjoen, G., Kaufman, J.-M. & Vincke, J. 2008. The role of testosterone in aggressive and non-aggressive risk-taking in adolescent boys. *Hormones and Behavior*, 53(3), 463-471.

Vermesh, M., Kletzky, O. A., Davajan, V. & Israel, R. 1987. Monitoring techniques to predict and detect ovulation. *Fertility and Sterility*, 47(2), 259-264.

Via, S. & Lande, R. 1985. Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution*, 39(3), 505-522.

Villemure, C. & Bushnell, M. C. 2007. The effects of the steroid androstadienone and pleasant odorants on the mood and pain perception of men and women. *European Journal of Pain*, 11(2), 181-191.

Vos, D. R. 1995. The role of sexual imprinting for sex recognition in zebra finches: a difference between males and females. *Animal Behaviour*, 50(3), 645.

Vukovic, J., Feinberg, D. R., Jones, B. C., DeBruine, L. M., Welling, L. L. M., Little, A. C. & Smith, F. G. in press. Self-rated attractiveness predicts individual differences in women's preferences for masculine men's voices. *Personality And Individual Differences*, In Press, Corrected Proof.

Walker, A. 1997. The Menstrual Cycle. London: Routledge.

Waynforth, D., Delwadia, S. & Camm, M. 2005. The influence of women's mating strategies on preference for masculine facial architecture. *Evolution and Human Behavior*, 26(5), 409.

Wedekind, C., Seebeck, T., Bettens, F. & Paepke, A. J. 1995. MHC-dependent mate preferences in humans. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 260(1359), 245-249.

Wedekind, C. & Furi, S. 1997. Body odour preferences in men and women: do they aim for specific MHC combinations or simply heterozygosity? *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 264(1387), 1471-1479.

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

Welling, L. L. M., Jones, B. C., DeBruine, L. M., Smith, F. G., Feinberg, D. R., Little, A. C. & Al-Dujaili, E. A. S. in press. Men report stronger attraction to femininity in women's faces when their testosterone levels are high. *Hormones and Behavior*, in press.

Wetzels, L. C., Hoogland, H. J. & de Haan, J. 1982. Basal body temperature as a method of ovulation detection: comparison with ultrasonographical findings. *Gynecologic and Obstetric Investigation*, 13(4), 235-240.

Wilcox, A. J., Weinberg, C. R. & Baird, D. D. 1995. Timing of sexual intercourse in relation to ovulation: Effects on the probability of conception, survival of the pregnancy and sex of the baby. *The New England Journal of Medicine*, 333(23), 1517-1521.

Williams, G. C. 1966. Adaptation and Natural Selection. Princeton: Princeton University Press.

Williams, J. M. & Dunlop, L. C. 1999. Pubertal timing and self-reported delinquency among male adolescents. *Journal of Adolescence*, 22, 157-171.

Willis, J. & Todorov, A. 2006. First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, 17(7), 592-598.

Wilson, D. S. 1994. Adaptive genetic variation and human evolutionary psychology. *Ethology and Sociobiology*, 15, 219 - 235.

Wilson, D. S., Clark, A. B., Coleman, K. & Dearstyne, T. 1994. Shyness and boldness in humans and other animals. *Trends In Ecology & Evolution*, 9(11), 442-446.

Wilson, D. S. 1998. Adaptive individual differences within single populations. *Philosophical Transactions of the Royal Society of London, Series B*, 353, 199-205.

Wilson, G. D. & Barrett, P. T. 1987. Parental characteristics and partner choice: some evidence for Oedipal imprinting. *Journal of Biosocial Science*, 13, 409-418.

Winkielman, P., Halberstadt, J., Fazendeiro, T. & Catty, S. 2006. Prototypes are attractive because they are easy on the mind. *Psychological Science*, 17(9), 799-806.

Wyart, C., Webster, W. W., Chen, J. H., Wilson, S. R., McClary, A., Khan, R. M. & Sobel, N. 2007. Smelling a single component of male sweat alters levels of cortisol in women. *Journal of Neuroscience*, 27(6), 1261-1265.

Wyatt, T. D. 2003. Pheromones and Animal Behaviour. Cambridge: Cambridge University Press.

Wysocki, C. J. & Preti, G. 2004. Facts, fallacies, fears, and frustrations with human pheromones. *The Anatomical Record Part A: Discoveries in Molecular, Cellular, and Evolutionary Biology*, 281A(1), 1201-1211.

Yamazaki, K., Beauchamp, G. K., Egorov, I. K., Bard, J., Thomas, L. & Boyse, E. A. 1983. Sensory distinction between H-2b and H-2bm1 mutant mice. *Proceedings of the National Academy of Sciences of the United States of America*, 80(18), 5685-5688.

Zahavi, A. 1975. Mate selection – a selection for a handicap. *Journal of Theoretical Biology*, 53, 205-214.

Zahavi, A. 1977. The cost of honesty. *Journal of Theoretical Biology*, 67, 603-605.

Zajonc, R. B. 1968. Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, 9, 1-27.

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.

Zebrowitz-McArthur, L. & Montepare, J. M. 1989. Contributions of a babyface and a childlike voice to impressions of moving and talking faces. *Journal of Nonverbal Behavior*, 13(3), 189-203.

Zei, G. P., Astolifi, P. & Jayakar, S. D. 1981. Correlation between father's age and husband's age: a case of imprinting. *Journal of Biosocial Science*, 13, 409-418.

Zeil, H. K. & Finkle, W. D. 1975. Increased risk of endometrial carcinoma among users of conjugated estrogens. *The New England Journal of Medicine*, 293, 1167-1170.

Zuckerman, M., Miyake, K. & Hodgins, H. S. 1991. Cross-channel effects of vocal and physical attractiveness and their implications for interpersonal perception. *Journal of Personality and Social Psychology*, 60(4), 545-554.

Zuckerman, M. & Hodgins, H. S. 1993. Developmental changes in the effects of the physical and vocal attractiveness stereotypes. *Journal of research in personality*, 27(4), 349-364.

Zuckerman, M., Hodgins, H. S. & Miyake, K. 1993. Precursors of interpersonal expectations: The vocal and physical attractiveness stereotypes. In: *Interpersonal Expectations: Theory, Research, and Applications* (Ed. by Blanck, P. D.), pp. 194-217. Cambridge: Cambridge University Press.