SEX DIFFERENCES AND THE ROLE OF SEX HORMONES IN FACE DEVELOPMENT AND FACE PROCESSING

Klára Marečková, MSc.

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

JULY 2013

ABSTRACT

Sex differences have been identified in both external appearance of faces (e.g. Bulygina et al., 2006; Weston et al., 2007) and the way information about faces is extracted by our brains, that is in face processing (e.g. Tahmasebi et al., 2012; Hampson et al., 2006). The mechanisms leading to the development of such sex differences are not well understood. This thesis explores the role of sex hormones in face development and face processing. Data from two large-scale studies (*Saguenay Youth Study* and *Imagen*, with n=1,000 and 2,000, respectively) and four smaller datasets (*Cycle-Pill Study*, n=20; *Pill Study*, n=20; *First Impression Study*, n=120, and *Twin Study*, n=119) were used to explore the effects of sex and sex hormones on face development (head MR images, MRI-face reconstruction) and face processing (functional MRI data, eye-tracking data).

Shape of male and female faces was influenced by both prenatal and pubertal androgens. Facial signature of prenatal androgens, identified by the sex-discordant twin design, was found also in an independent dataset of female adolescents (singletons) and we showed that prenatal androgens, indexed indirectly by the facial signature, were associated with larger brain size. We propose that this facial signature might be used, similarly to digit ratio, as an indirect index of prenatal androgens.

Variability in postnatal sex hormones due to the use of oral contraception and the phase of menstrual cycle influenced brain

response to faces. Using the same dynamic face stimuli as in the functional magnetic resonance imaging (fMRI) study, we showed that eye-movements scanning the face did not differ between the users and non-users of oral contraception.

We conclude that effects of sex hormones can be observed in both the face and the brain and that these effects help us understand sex differences in face shape and face processing.

Chapters 2-5 can be also accessed as the following four papers:

Marečková, K., Weinbrand, Z., Chakravarty, M.M., Lawrence, C., Aleong, R., Leonard, G., Perron, M., Pike, G.B., Richer, L., Veillette, S., Pausova, Z., & Paus, T. (2011). Testosterone-mediated sex differences in the face shape during adolescence: Subjective impressions and objective features. *Hormones and Behavior, 60*, 681-690.

Marečková, K., Chakravarty, M.M., Huang, M., Lawrence, C., Leonard, G., Perron, M., Pike, B.G., Richer, L., Veilette, S., Pausova, Z., & Paus, T. (2013). Does skull shape mediate the relationship between objective features and subjective impressions about the face? *Neuroimage, 79*, 234-240.

Marečková, K., Chakravarty, M.M., Lawrence, C., Leonard, G., Perruse, D., Perron, M., Pike, B.G., Richer, L., Veilette, S., Pausova, Z., & Paus, T. Identifying craniofacial features associated with prenatal exposure to androgens and testing their relationship with brain development. (in preparation for submission to *Journal of Neuroscience*).

Marečková, K., Perrin, J.S., Nawaz-Khan, I., Lawrence, C., Dickie, E., McQuiggan, D.A., Paus, T., & the IMAGEN Consortium (2012). Hormonal contraceptives, menstrual cycle and brain response to faces, *Social Cognitive and Affective Neuroscience.* PMID: 23175677

ACKNOWLEDGEMENTS

First of all, I would like to thank Prof. Tomáš Paus, with whom I started my MSc program in Cognitive Neuroscience and Neuroimaging at the University of Nottingham in September 2008 and with whom I have moved to Toronto in January 2010 to study for my PhD at the Rotman Research Institute at Baycrest. I owe him all I have learned in neuroscience and research. I would also like to thank Dr. Claire Lawrence, my co-supervisor at the University of Nottingham, who has made the co-supervision arrangements possible. Our monthly skype meetings have been a great source of motivation and advice.

I would also like to thank all my collaborators from the Paus lab and beyond. In particular, I would like to thank all those, who have helped to design and collect data for Saguenay Youth Study, Imagen, Twin study, and Cycle Pill study. The datasets are invaluable. Working with Dr. Mallar Chakravarty on the face project was inspiring and showed me what all is possible with image processing. Dr. Nick Qiu taught me much about scripting. Dr. Erin Dickie helped me to regain confidence after receiving feedback from the reviewers of my manuscript. I would also like to thank Mei Huang and Zohar Weinbrand, undergraduate students from University of Toronto, who helped me to recruit participants for the First Impression Study from the UofT psychology pool. Being part of the lab has been a great learning experience!

I would also like to thank number of great institutions, who triggered my learning and will remain close to my heart, ever. Rotman Research Institute at Baycrest hospital, who organized many talks, workshops, and let me use their eye-tracker and blood lab facilities. In particular, I would like to thank Dough McQuiggan who taught me much about eye-tracking. The School of Psychology and the Medical School at University of Toronto let me audit their graduate courses. The courses on Univariate Statistics, Advanced Statistics, Neuroanatomy, Sex and the Brain, and Imaging Genetics have all been tremendously useful. I am also grateful I have attended the Annual Meeting of the Organization for Human Brain Mapping in Quebec City in 2011, which has been a great source of motivation at an early stage of my PhD studies, especially for the replication of the effects of oral contraception use on brain response to faces and studying the effect of oral contraception duration. I am grateful also to all the funding agencies that have supported me throughout.

Last but not least, I would like to thank my boyfriend, Milan Linkesch, who moved with me to Canada, my brother, Jakub Marecek, who gave me confidence to pursue the PhD, and all my friends who made me happy, energized, and part of the Toronto community, including my dressage coach Abbey Simbrow, Community Association for Riding for the Disabled, and Toronto Outdoor Club.

Thank you.

CONTENT

Face is a land one can never tire of exploring

(Dreyer, 1955).

CHAPTER 1

Introduction

1.1 FACE AS A SOURCE OF INFORMATION

Face is a rich source of information. Features of the face signal one's sex, age, and ethnicity, reveal information about one's physical (e.g. malnutrition) and mental (e.g., depression) health, and allow inferences about behavioral dispositions (Carre, & McComrick, 2008). For example, increase in facial width-to-height ratio was associated with low perceived trustworthiness (Stirrat, & Perrett, 2010), higher dominance, and aggressive behavior (Carre, & McComrick, 2008). Further research showed that judgments of dominance predicted career success (Mueller & Mazur, 1996). Our ability to read facial cues allows us to decide whether a person might be a potential mate or competitor and is thus essential for appropriate social interactions.

Faces catch our attention more than any other visual object (Bindemann et al., 2005). While approximately 200 ms are required to recognize an object, only 100 ms is sufficient to discriminate a face (Bentin et al., 1996). Our fine-tuned expertise in face perception can be measured also at the level of the brain. Kanwisher et al. (1999) showed

that fusiform face area (FFA), brain region located in the fusiform gyrus of the temporal lobe, responds more to faces than other objects. Lesions in this area are characteristic for people with prosopagnosia, the “face-blindness” disease (Barton et al., 2002; Farah, 1990). Further research described also other parts of the brain involved in face processing (Posamentier & Abdi, 2003; Haxby et al., 2002). While FFA seems to be involved mainly in face perception, superior temporal sulcus (STS) is involved in more dynamic aspects such as eye and lip movements or gaze direction (Haxby et al., 2002; Engel & Haxby, 2007), and amygdala evaluates valence of the face and by interaction with other regions influences approach/avoidance decisions (Todorov, & Engell, 2008).

1.2 THE IMPORTANCE OF FACES IN SOCIAL INTERACTIONS

Our sensitivity to facial cues and ability to interpret them is essential for correct recognition of emotion (Pierce, 2012). The ability to read faces effectively gives one an advantage in social interactions (Schultz, 2005). Impaired sensitivity to facial cues often leads to problems with emotions and approach-avoidance behavior, which are the symptoms of many psychiatric conditions (Lombardo et al., 2012a). For example, autistic individuals struggle with perception of gaze direction (Baron-Cohen et al., 1997) and facial affect (Bormann-Kischkel et al, 1995), and have diminished rates of eye contact (Hobson, & Lee, 1998) and social interaction with others (Pierce, & Schreibman, 1995). Pierce et al. (2001) suggested that this limited

experience with faces most likely triggers a whole cascade of events that finally result in severe problems with social communication.

1.3 SEX DIFFERENCES IN FACE SHAPE

Sex differences in face shape have been described at both the level of facial tissue (Carre & McCormick, 2008; Penton-Voak et al., 2001; Toma et al., 2008) and the skull (Weston et al., 2007; Bulygina et al., 2006). Males (vs. females) have broader and shorter upper face (distance between the lip and the brow) than expected for its length and this sex difference was observed in both humans (Weston et al., 2007) and chimpanzees (Weston et al., 2004). While most authors seem to report sex differences in face shape as of puberty (Weston et al., 2007; Ferrario et al., 1998; Enlow et al., 1996), Bulygina et al. (2006) were able to detect sex differences in skull features already in 6-month old human infants: male infants had smaller faces but more globular frontal bones than female infants (Bulygina et al., 2006). This sex difference reversed and became more evident at approximately the age of 12-14 when male faces became larger and had smaller and flatter frontal bone than female faces (Bulygina et al., 2006).

1.4 SEX DIFFERENCES IN FACE PERCEPTION

Sex differences have been observed also in face perception. Women were better than men in face detection, the very first stage of face perception (Cohen's d=0.91; McBain et al., 2009). Women also showed a consistent advantage in recognition of emotions in the face

(Hampson et al., 2006; Hoffmann et al., 2010). Sex differences in reaction time during an emotion recognition task were particularly amplified for negative emotion recognition (Hampson et al., 2006). Emotions displayed at mid-intensities were accurately recognized by women but not men (Hoffmann et al, 2010). Hall et al. (2010) showed that the female advantage in emotion recognition can be predicted from dwell time and number of fixations to the eyes when scanning the face. Meta-analytic review about sex differences in emotion recognition reported presence of sex differences as early as infancy and showed that these persist throughout childhood and adolescence (McClure, 2000). McClure (2000) suggested they might be related to development of neural systems important for emotion processing. Female advantage in emotion recognition was seen also in men and women suffering from psychiatric disorders. Scholten et al. (2005) showed that women diagnosed with schizophrenia were more accurate in emotion recognition than men diagnosed with schizophrenia and suggested that this might explain why female patients are less impaired in social life than male patients.

1.5 DO SEX HORMONES CONTRIBUTE TO THE DEVELOPMENT OF SEX DIFFERENCES IN FACE SHAPE?

Sex hormones create an internal environment that influences target tissues (e.g., face and brain tisues). This thesis will focus on the role of sex hormones as one of the possible mechanisms underlying sex differences in face shape and face perception. Sex hormones

shape sexual differentiation of both the body (e,g, Lazic et al., 2011) and brain (e.g. Lombardo et al., 2012b) since prenatal stage. Exposure to androgens and estrogens as well as the number of their receptors differ between male and female fetuses. While female fetuses are exposed only to very small amounts of prenatal androgens produced in the adrenal gland, male fetuses develop testes and are exposed to much higher levels of prenatal androgens.

The appearance of sex differences in face shape during adolescence (Bulygina et al., 2006; Weston et al., 2007) suggests that sex hormones might play a role. Sex hormones influence growth of bones in humans (e.g. Morishima et al., 1995). Animal experiments showed that orchiectomy and ovariectomy induced bone loss and that estrogens and androgens prevent bone loss during adolescence (Fujita et al., 2001). Low doses of testosterone treatment triggered craniofacial growth in boys with delayed puberty and the effects of testosterone were most pronounced in the mandible, cranial base, and anterior face height (Verdonck et al., 1999). Testosterone levels were also related to masculine facial structure (Pound et al., 2009), and subjective impressions about masculinity of a face (Penton-Voak, & Chen, 2004). This thesis will further explore the role of androgens in the development of sex differences in face shape and consider the possibly different effects of prenatal and postnatal sex hormones, as suggested by the organizational and activational hypothesis (Phoenix et al., 1959; see section 1.4. for more information).

1.6 DO SEX HORMONES CONTRIBUTE TO THE DEVELOPMENT OF SEX DIFFERENCES IN FACE PERCEPTION?

It seems that sex hormones might also contribute to the development of sex differences in face perception during both pre-natal and post-natal periods. Previous research showed that foetal testosterone measured from amniotic fluid was negatively related to infant's eye contact (Lutchmaya et al., 2004), and children's (6-9 years old) ability to discriminate other's emotional facial expresions (Chapman et al., 2006). Neither Van Honk et al. (2011) nor Voracek & Dressler (2006), however, found a relationship between ability to discriminate others' emotional facial expressions (using the same task as Chapman et al., 2006) and digit ratio, the indirect index of prenatal androgens, in adults. While the presence of this relationship in children and absence of this relationship in adults might be related to possibly low reliability of digit ratio as an index of prenatal androgens, it is also likely that it might be related to the appearance of postnatal hormones during puberty.

The effect of postnatal sex hormones has been well described in other domains of cognition: estrogen was associated with better performance on verbal tasks and worse performance on spatial tasks in women (e.g. Maki et al., 2002; Kimura, 1999). Only a handful of studies have explored the possible effect of ovarian sex hormones on emotion recognition in the face in particular. Derntl et al. (2008) showed that women in the follicular phase of menstrual cycle were better in recognition of emotions in the face than women in the lutheal phase and

that levels of progesterone negatively correlated with performance on emotion recognition task. Pearson and Lewis (2005) also reported a relationship between ovarian hormones and recognition of emotions in the face, but their findings showed better fear recognition during pre-ovulatory phase compared with menstruation, suggesting that it is rather estrogen than progesterone that influences recognition of emotions in the face. This thesis will explore the possible role of ovarian sex hormones on brain response to faces and eye-movements scanning the face.

1.7 THE ROLE OF PRENATAL VERSUS PUBERTAL SEX HORMONES IN SEX DIFFERENCES

Early development and puberty are two stages of life when sex hormones influence our bodies, brains, and behavior. Early development is characterized by exposure to prenatal androgens. Puberty is characterized by an increase in testosterone, the onset of menarche and – since the 1960's - the possible use of oral contraception in girls. The organizational and activational hypothesis (Phoenix et al., 1959), describes differences in the actions of sex hormones during these two stages of life. The prenatal period is seen as a critical window when androgens organized and permanently masculinized the organism (Phoenix et al., 1959). Hormones that appeared later in life (e.g. puberty) activated target organs and mating but their effects were reversible and acute (Phoenix et al., 1959). Female guinea pigs whose mothers were treated with testosterone

during pregnancy showed permanent reduced lordosis behavior and permanent masculinisation of external genitalia (Phoenix et al., 1959). Postnataly, these females were also more responsive to testosterone than control females (Phoenix et al., 1959). Amounts of testosterone which were effective prenatally had no similar lasting effects when administered postnatally (Phoenix et al., 1959).

The organizational and activational hypothesis is an old, but still prevalent theory. More than 50 years of testing extended its original focus on sex behavior (Phoenix et al., 1959) to other types of behavior (e.g. Meaney & Stewart, 1981), and brain as their intermediate phenotype. Sex differences were found not only in the brain nuclei relevant for reproduction (Murakami & Arai, 1989; Davis et al., 1996), but across the whole brain. Gonadal hormones regulated the axon, dendrites (Torran-Allerand, 1976), and synaptic differences (Parducz et al., 2002). Twin studies describing effects of prenatal androgens on brain structure (e.g. Peper et al., 2009) and function (e.g. Cohen-Bendahan et al., 2004) triggered interest in the effect of co-twin's sex (and thus prenatal androgens) on other phenotypes such as cognition (e.g. Galsworthy et al., 2000; Vuoksimaa et al., 2010), or susceptibility to diseases (Culbert et al., 2008; Ho et al., 2005). Prenatal androgens, measured from amniotic fluid, were found to have a negative effect on empathy in both boys and girls (Chapman et al., 2006), mentalizing at the age of two (Knickmeyer et al., 2006), and quality of social relationships (Knickmeyer et al., 2005). Studies on females with cogenital adrenal hyperplasia (CAH), who are by definition exposed to

higher levels of androgens, showed reduced empathy (Mathews et al., 2009) and increased physical aggression (Pasterski et al., 2007). Indeed, Baron-Cohen (2002) has suggested that the exposure to prenatal androgens influences the risk of autism.

More than 50 years of testing did not seriously question the principles of organizational and activational hypothesis (Arnold, 2009) and only two alternative hypotheses appeared – aromatization hypothesis (Naftolin et al., 1975) and the extended critical window hypothesis (Schultz, 2009). The aromatization hypothesis explained why administration of exogenous estradiol produced similar, or even larger, masculinising effects as administration of testosterone. Naftolin et al. (1975) used a rodent model to show that aromatase converts testosterone to estradiol, which then binds to estrogen receptors in the critical regions of neonatal brain (Naftolin et al., 1975). Further research showed, however, that aromatization hypothesis could not account for all sex differences in brain morphology and behavior (reviewed in Zuloaga, 2008). In primates, androgens seem to act directly on the androgen receptor to masculinise the brain (Zuloaga, 2008). Wallen and Baum (2002) concluded that aromatization is more important for male sexual differentiation of altricial species (e.g. rats, mice, ferrets for whom smaller portion of brain development occurs in utero) than precocial species (e.g. guinea pig, pig, monkey, human for whom greater portion of brain development occurs in utero).

The possibility of an extended critical window was proposed by Sisk and Zehr (2005) who suggested that secretion of gonadal hormones during puberty might organize further the adolescent brain and thus provide a second critical window for organization of the neural circuits. Pubertal hormones were shown to be responsible for the enlargement of locus coeruleus in females (vs. males; Pinos et al., 2001) and the enlargement of primary visual cortex in males (vs. females; Nunez et al., 2002). The amygdala, hippocampus, and bed nucleus of stria terminalis also developed differently in males and females during adolescence (Sisk & Zehr, 2005). White matter volume increased in adolescence, but in adolescent males in particular (Giedd et al., 1999; Perrin et al., 2009). It seems that pubertal hormones first *organize* neural circuits in the developing adolescent brain and then these long-lasting structural changes determine behavioural response to hormones and socially relevant sensory stimuli (Schultz et al., 2009). According to the extended critical window hypothesis, sensitivity to the organizing actions of testosterone gradually decreases and developmental sensitivity to hormones and organization of the brain and behavior seems to be terminated by the end of adolescence (Schultz et al., 2009).

While experimental testing of these three hypotheses suggested by Phoenix et al. (1959), Naftolin et al. (1975), and Schultz et al. (2009) respectively would be an interesting area for animal research, ethical reasons do not allow such experimental testing in humans. Longitudinal twin dataset would have an ideal design to explore these three

paradigms in humans, but access to such datasets is sparse. We will use these three paradigms to generate questions and predictions about the effects of sex hormones on face development and face processing. This thesis will (1) explore the influence of prenatal and pubertal androgens on face development, (2) study the effect of face shape on impressions about the face, and (3) clarify the role of sex hormones in brain response to faces.

1.8 MEASUREMENT OF PRENATAL SEX HORMONES

Prenatal androgens can be measured directly in (a) amniotic fluid, (b) umbilical cord blood, or (c) through maternal testosterone levels during pregnancy; none of these data are readily available for human samples. Study designs using (a) opposite-sex dizygotic twins, (b) individuals with disorders of sexual development, or (c) individuals who have been exposed to chemicals that mimic or block endogenous hormones are thus used to compare individuals with (presumed) higher vs. lower exposure to prenatal androgens. Since these data are also hard to access, some studies used the digit ratio (e.g. Manning et al., 1998), an indirect measure of prenatal androgens. Its reliability, however, has been questioned (e.g. Berenbaum et al., 2009).

This thesis will (i) use twin design to explore the effect of prenatal androgens on face shape and (ii) try to derive a new indirect index of prenatal androgens. Literature on sex differences in face shape suggests that prenatal androgens might have left their signature in the face: Sex differences in craniofacial morphology were described in 6-

month old human infants (Bulygina et al., 2006). Studies in adults showed a relationship between digit ratio and perceived masculinity of the face (Neave et al., 2003). The size of teeth differed between androgenized female monkeys and female controls (Zingeser, & Phoenix, 1978), female-to-male transsexuals and female controls (Antoszewski et al., 2009), as well as females with a male and female co-twin (Dempsey et al., 1999). We will use twin design and head MR images to explore the possiblity of finding a new and hopefully more reliable indirect index of prenatal androgens in the face. If successful, we will use this facial signature to study the effects of prenatal androgens on brain.

1.9 MEASUREMENT OF POSTNATAL SEX HORMONES

Postnatal sex hormones can be measured easily in blood (plasma or serum), saliva, or urine. Only a fraction of the measured sex hormones is bioavailable and not bound to sex hormone binding globulin (SHBG). For example, testosterone is present in both protein-bound (testosterone bound to SHBG, testosterone bound to albumin) and non-protein bound (free testosterone) form but only the free and albumin bound testosterone can be absorbed by tissues and are thus called bioavailable testosterone. Therefore, measurement of SHBG has to accompany measurements of sex hormones (from blood) in order to calculate levels of bioavailable androgens and estrogens.

Timing of the sampling of sex hormones is essential. Levels of testosterone vary during the day and reach their peak in the morning

(Ankarberg-Lindgren, & Norjavaara, 2004). Levels of estrogen and progesterone vary as a function of menstrual cycle (Hampson & Young, 2007): while levels of estrogen increase during the follicular phase, peak during mid-cycle, and decrease during luteal phase, levels of progesterone start to increase at mid-cycle and peak during luteal phase. Menstruation phase is characterized by low levels of both estrogen and progesterone. Levels of sex hormones in females might be further influenced by use of oral contraception, which contains exogenous estrogens (usually ethinyl-estradiol) and progestins and reduces levels of endogenous estrogen, progesterone (Hampson & Young, 2007), and testosterone (Graham et al., 2007; Hietala et al., 2007).

We collected morning blood samples to measure serum levels of testosterone (see Chapter 2 for details), estrogen, and progesterone (see Chapter 5 for details). Levels of bioavailable testosterone were calculated from the total testosterone and sex-hormone-binding globulin using the Sodergard et al (1982) formula and related to face shape in both males and females. In women, we also collected information about menstrual cycle phase and oral contraception use and explored their effects on brain response to faces and eye-movements scanning the face.

1.10 THESIS DESIGN

My research started as a follow-up of the Tahamsebi et al (2012) findings about sex differences in face perception and explored the role

of postnatal sex hormones on brain response to faces. Subsequently, I learned about the organizational and activational hypothesis and was trying to find an accessible index of exposure to prenatal testosterone that could be used to study the relationship between prenatal testosterone and the brain. Studies about sex differences and the role of sex hormones in craniofacial development were thus triggered by the search for such an indirect index of prenatal testosterone.

The thesis as such thus presents effects of sex and sex hormones in two main areas: the observer – and the observed . The first part of the thesis focused on the face of the observed individual and described (i) sex differences in face shape and (ii) the role of prenatal and postnatal sex hormones in face development. We also studied how the sex hormone-related features of the face contribute to correct sex identification of the face by an observer. The second part of the thesis focused on the observer of the face and explored the effects of sex hormones on (i) eye-movements when scanning faces and (ii) brain response to faces.

1.11 THE OUTLINE OF THE THESIS

Sex differences and the effects of sex hormones will be explored at the level of the face shape (Chapter 2), skull shape (Chapter 3), brain size (Chapter 4), and brain response to faces (Chapter 5). Chapters 2 and 5 will focus on the effects of postnatal sex hormones on the face

shape and the brain response to faces, respectively. Chapter 4 reports two studies which describe the effects of prenatal sex hormones on cariofacial features (Study 1) and brain size (Study 2). Chapter 3 will investigate the relationship between the tissue-related (Chapter 2) and skull-related (Chapter 4) features of the face.

1.12 METHODOLOGICAL ASPECTS OF THE THESIS

Variability in facial features (Chapter 2, 3, and 4) were studied using head MR images. We placed landmarks on a) the skull and b) facial tissue of head MR images and then explored group differences in the location of these landmarks. This was done in two ways: (1) calculating euclidean distances between landmarks, (2) using principal component analysis (PCA) to extract the main features describing variability in the landmarks. This methodology allowed us to study the face shape in 3D, stripped from any additional external characteristics such as hair cut, make-up, or shape of eye-brows, and also differentiate between the effect of tissue (muscle and fat) and the skull.

Perceptual “sampling” of a face was examined using eye tracking techniques (Chapter 5). Observers were presented with videoclips of faces, and the length and number of fixations in four main areas of each face (eyes, nose, mouth, and the rest of the face) was calculated.

Brain response to faces (also Chapter 5) was studied using functional magnetic resonance imaging (fMRI). Participants were presented with identical facial videoclips as used for the eye-tracking

study and blood-oxygen-level-dependent (BOLD) response to these stimuli was calculated in two ways: (1) voxel-wise, using the whole brain field of view, (2) focussing on the fusiform face area, which is the main region of interest.

Subjective impressions about the sex of a face (Chapter 2 and 3) were assessed by presenting observers with MRI-reconstructed faces and asking them to rate the sex of the individual.

1.13 DATASETS USED IN THE THESIS

The role of sex and sex hormones in face development and face processing was studied using available structural and functional MRI data, as well as other data, obtained in both large-scale and small-scale datasets. These included two large scale studies (i) *Saguenay Youth Study*, a study of 1,000 Canadian adolescents (aged 12-18), (ii) the *Imagen*, a study of 2,000+ European adolescents (age 14 years old), and two smaller scale studies (i) a study of 20 young women who are either taking oral contraception or freely cycling (aged 18-29), and (ii) a study of 119 twins (8-year old). While the *Saguenay Youth Study*, the *Imagen*, and the *Twin study* were not designed to answer the particular questions of this thesis, they were applicable and available to do so and we used novel ways of data analysis to test our hypotheses. In addition, an eye-tracking study of 20 women aged 18-29 (Chapter 5) and a face perception study of 120 women aged 17-30 (Chapter 3) were designed specifically for this dissertation. These small-scale experimental

datasets are introduced in Chapters 5 and 3 respectively, but the two large scale datasets are described below.

1.13.1 Saguenay Youth Study (SYS)

The Saguenay Youth Study (Pausova et al., 2007) is a study of 12-18-year-old adolescents (n=1,024) of French-Canadian origin from the Saguenay Lac-Saint-Jean region in Quebec, Canada. The main aim of the study was to investigate long-term concequences of prenatal exposure to maternal cigarette smoking. Adolescents who were not exposed prenatally to maternal cigarette smoking were matched with exposed adolescents by their school and level of maternal education and assessed for the following phenotypes: (1) brain, abdominal fat, and kidney MRI, (2) cardiovascular, body-composition, and metabolic assessment, (3) cognitive assessments, (4) questionnaires about life habits, personality, and psychiatric symptoms. Chapter 2 used 597 MRIs (292 males, 305 females) that were available for MRI-face reconstruction at that point of SYS data collection. Experiments presented in Chapter 3 were conducted later, and therefore we could use 876 MRIs (411 males, 462 females; information about the sex of 3 pariticipants was missing) that were available for MRI-face reconstruction. MRI data from 462 females were also used in Chapter 4 (Study 2) for testing the relationship between the facial signature of prenatal androgens and brain size.

1.13.2 Imagen

The Imagen study is a study of 2,000 14-year-old adolescents from Europe (Germany, England, Ireland, France) across eight acquisition sites, investigating mental health and behavior in teenagers. Adolescents took part in an 8-hour long testing including (1) brain MRI, (2) blood sampling for genetic analyses, (3) cognitive testing, (4) behavioral assessment using questionnaires and structured interview. A total of 55 adolescent girls from this dataset were using oral contraception. We matched these 55 girls by age, pubertal stage, and acquisition site with 55 freely cycling girls and used this sample in Chapter 5 (Experiment II), as a replication of our findings from Chapter 5, Experiment I.

1.14 CHAPTERS OVERVIEW: MAIN RESEARCH QUESTIONS

1.14.1 Chapter 2 – The effect of sex hormones on face development

The development of sex-specific facial features was described in two ways: (1) objectively, using landmarks on the face and principal components that explain variability in facial features; and (2) subjectively, using ratings about the sex of the face as perceived by female undergraduates. Testosterone and body fat were identified as main predictors of face shape in both males and females. While the face kept developing during adolescence in males, it seemed to be fully developed in females at the age of 12. Females with high loadings of

maleness in the face and thus often miss-classified as males were followed up and compared with correctly classified females with low loadings of maleness in the face. We have already published results of this investigation in *Hormones and Behavior* under the following title: “Testosterone-mediated differences in the face shape during adolescence: Subjective impressions and objective features” (see Appendix 1).

1.14.2 Chapter 3 – Does skull shape mediate the relationship between objective features and subjective impressions about the face?

The findings from Chapter 2 did not give the answer as to whether it is skull or facial tissue that influences sex identification of the face. Chapter 3 examines this directly to assess whether skull features mediate the relationship between objective facial features and subjective impressions about the sex of the face. Bootstrapping and mediation analysis showed that skull mediated the relationship between objective facial features and subjective impressions about male but not female faces. Skull revealed as a mediator of the relationship between objective facial features and subjective impressions about the female face only after the facial features were adjusted for body fat. While body fat had a slight positive effect on correct sex recognition of male faces, there was a negative effect of body fat on correct sex recognition of female faces and craniofacial bone structure alone could not explain the relationship between facial features and identification of a face as

female. We have already published these findings in *Neuroimage* under the following title: "Does skull shape mediate the relationship between objective features and subjective impressions about the face?“ (see Appendix 2).

1.14.3 Chapter 4 – Signature of prenatal androgens in the face

Chapter 4 uses a twin design and a morphometic analysis of head MR images to identify a signature of prenatal androgens in the face that could possibly complement the only readily available index of prenatal adrogen exposure, namely the digit ratio. Females with a female co-twin showed facial features that distinguished them from all other twin groups (OSF, OSM, SSM) exposed to at least some levels of prenatal androgens. The effect size of all three comparisons was large. In order to verify the existence of the relationship between prenatal androgens and facial features, we studied relationship of this facial signature with brain size, a known correlate of prenatal androgens, in a large independent sample of adolescent females. Facial signature could explain 2% and the mean distance between the sides of the jaw and chin even 8% of variance in brain size. We propose that this facial signature might be used as an indirect index of exposure to prenatal androgens, expecially by researchers who have access to T1-weighted head MRI but not direct measures of prenatal androgens from amniotic fluid or umbilical cord blood. These findings are in preparation for submission to *Journal of Neuroscience* under the following title:

"Identifying craniofacial features associated with prenatal exposure to androgens and testing their relationship with brain development “.

1.14.4 Chapter 5 – The effect of sex hormones on face processing

Chapter 5 builds on the literature reporting sex differences in face perception that shows a consistent female advantage (e.g. McBain et al., 2009; Hampson et al., 2006; Hall & Matsumoto, 2004; Tahmasebi et al., 2011) and explores the effect of sex hormones on brain response to faces. Phase of menstrual cycle and use of oral contraception are studied as the predictors of BOLD response in the face processing network of young adult women. Effects of oral contraceptives are replicated in a sample of female adolescents and followed up by an eye-tracking study exploring whether the increased brain response is reflected by a particular face-scanning pattern. We have already published results of this investigation in *Social Cognitive and Affective Neuroscience* under the following title: “Hormonal contraceptives, menstrual cycle and brain response to faces” (see Appendix 3).

CHAPTER 2

Testosterone-mediated sex differences in the face shape during adolescence: subjective impressions and objective features

2.1 INTRODUCTION

Faces are highly informative: even a simple silhouette of a face, devoid of any specific cues, can allow an observer to identify its sex, age or race (e.g., Davidenko, 2007; Martin & Macrae, 2007). Correct identification of sex from a face is important as it assists the perceiver in deciding whether a person may be a potential mate or competitor. The degree of the masculinity of a given face has been described as indicator of dominance and health (Perret et al., 1998). Fink et al. (2005) described an association between the face masculinity/dominance and high testosterone-to-estrogen ratio (T/E ratio). The latter was also associated with larger cheekbones, mandible and chin, lengthening of the lower face and a forward prominence of the eyebrow ridges. On the other hand, low T/E ratio was associated with a more gracile face characterized by smaller mandible, fuller lips and high eyebrows (Fink et al., 2005).

Several studies identified sexual dimorphisms in various craniofacial phenotypes, both when considering only the bone tissue,

namely the skull (e.g., cephalometry: Weston et al., 2007; X-rays: Bulygina et al., 2006) or when measuring the face surface and, therefore, including also the soft tissue covering the skull (e.g., photographs: Carré & McCormick, 2008, Penton-Voak et al. 2001; laser scanning: Toma et al., 2008). Specific sex differences in facial shape and when they arise developmentally are not well understood, however. Using X-rays of the face and cranium obtained from 14 males and 14 females aged between 1 month and 21 years of age, Bulygina et al., (2006) observed sexually dimorphic features as early as 6 months of age: male infants had smaller faces than females, with more globular frontal bones. Between 12 and 14 years of age, this difference reversed such that male faces were larger than those of females, with smaller and flatter frontal bones. Weston et al. (2007) studied 68 male and 53 female skulls aged between 9 months of age to 30 years of age and observed that sexual dimorphism was first evident between 12 and 14 years of age. Longitudinal studies of craniofacial structure and development (Bulygina et al., 2006; Thordarson et al., 2006) suggest, therefore, that a distinct set of differences emerges during puberty. If sex differences in craniofacial development are mediated in part by sex hormones, once sex hormone secretion increases at the onset of puberty, craniofacial structure may exhibit corresponding changes. This is consistent with the reports of sexual domorphism in craniofacial structure arising in early (12 to 14 years) puberty (Weston et al., 2007; Bulygina et al., 2006).

In the present study, we investigated sex differences in facial development in typically developing, healthy adolescents aged between 12 and 18 years from the Saguenay Youth Study (SYS; Pausova et al., 2007). Using magnetic resonance images (MRI) of the adolescents' heads acquired from the SYS (Pausova et al., 2007), our group has recently developed a novel computational method for analyzing facial features (Chakravarty et al., 2011). In the present study, we sought to determine when sex-related differences become perceptually apparent to observers. Raters were asked to identify the sex of the computed MRI-reconstructed face images from the SYS dataset. The difference or correspondence between objective sex and raters' judgements of sex were then used to explore which facial characteristics improved classification of facial sex and enabled a perceptual signature of sex differences in the face to be described. We also used levels of bioavailable testosterone and a genetic polymorphism in the androgen receptor gene (*AR*) to evaluate the effects of testosterone on the development of these sex differences during puberty.

2.2 MATERIALS AND METHODS

2.2.1 Raters

Eighty-eight first-year undergraduate students of psychology (28 males, 60 females) from the University of Toronto Mississauga (Toronto, Ontario, Canada) were recruited to rate the sex of MRI-reconstructed faces. The mean age of male raters was 19.4 years (SD

= 1.79; age range 17-26 years), and the mean age of female raters was 18.7 years (SD = 1.21; age range 17-23 years). Students participated for course credit and a 1-in-6 chance to win a gift voucher for $50 CAD. Recruitment criteria required all raters to be of White Caucasian ethnicity to match the White Caucasian background of the MRI-reconstructed faces of the SYS adolescents they would be rating. These criteria eliminated potential confounding effects by the other-race effect (Malpass & Kravitz, 1969; Bothwell et al., 1989). No rater was taking antidepressant or antipsychotic medication at the time of their participation in this study.

2.2.2 Source sample

The Saguenay Youth Study (SYS; Pausova et al., 2007), a large study of adolescents, was used as a source sample. The SYS includes (1) MRIs of brain, abdominal fat, and kidneys, (2) standardized and computer-based neuropsychological tests, (3) hospital-based cardiovascular, body-composition and metabolic assessments, and (4) questionnaire derived measures about personality, psychiatric symptoms, drug and alcohol use, and life habits (Pausova et al., 2007). In this study, we use puberty development scales, levels of bioavailable testosterone, androgen receptor genotype, and T1-weighted magnetic resonance (MR) images from a sample of these typically developing adolescents (n = 597; 292 male, 305 female, age range = 12 to 18 years). T1-weighted MR images provide very good contrast between

soft tissues and therefore could be used not only for the analysis of brain anatomy but also for the analysis of craniofacial structure.

2.2.2.1 Puberty development

Puberty Development Scale (Peterson et al., 1988), an 8-item self-report measure of physical development, was used to determine a Tanner stage of each participant. This self-report measure correlates with physician ratings of pubertal development (Dorn et al., 1990). Separate forms for males and females include questions about growth in stature, pubic hair, menarche in females and voice changes in males and enable to categorize each adolescent into one of the following five Tanner stages: (1) prepubertal, (2) beginning pubertal, (3) midpubertal, (4) advanced pubertal, and (5) postpubertal.

2.2.2.2 Bioavailable testosterone

Serum testosterone levels were determined as previously described by Perrin et al. (2008). Fasting blood samples were collected between 8:00 A.M. and 9:00 A.M. and radioimmunoassay analysis (Testosterone RIA DSL-4000; Diagnostic Systems Laboratory) was performed. Levels of bioavailable testosterone were calculated from the serum testosterone and sex-hormone-binding-globulin using an equation formulated by Södergård et al. (1982).

2.2.2.3 Androgen receptor genotype

We determined the number of CAG repeats in Exon 1 of the androgen receptor gene (AR) as described previously (Perrin et al., 2008). *AR*s with longer polyglutamine stretches encoded by greater numbers of CAG repeats in Exon 1 of the *AR* gene appear to be associated with a lower transcription activity of *AR* (Irvine et al., 2000). In this way, *AR*s encoded by genes with greater CAG repeats are predicted to exert less optimal effects by *AR* ligands, such as testosterone. Briefly, PCRs were performed using 100 ng of genomic DNA in 8.0 µL volume of 1.0 mM $MgCl_2$ (Qiagen), 1x PCR buffer containing 1.5 mM $MgCl_2$ (Qiagen), 0.035 µM dNTPs (Qiagen), 0.04 U/µL HotstarTaq DNA polymerase (Qiagen), and 200 nM forward and reverse primer. Samples were denatured at 95ºC for 10 min, followed by 45 cycles containing a denaturing phase at 95ºC for 30s, an annealing phase at 60ºC for 30 s, and an extension phase at 72ºC for 30 s. The final extension phase was performed at 72ºC for 7 min. Using 2 µL of PCR products, 0.15 µL of Genescan 500 Liz size standard (Applied Biosystems), and 8.5 µL of Hi-Di Formamide (Applied Biosystems), a reading mixture was prepared and migrated on Applied Biosystems 3730xl DNA Analyzer. The applied biosystems GeneMapper analysis program (release 3.7, October 12, 2004) was used to analyse the genotypes.

The male adolescents studied in this report possessed CAG repeat lengths that ranged between 8 and 32 CAG repeats. The median

was 21.7 repeats, and this was used to split the adolescents into males with the “short” *AR* (<22 CAG repeats), and those with a “long” *AR* gene (≥22 CAG repeats).

2.2.2.4 Deriving faces from MRI

Analysis of craniofacial structure was performed in the fashion described by Chakravarty et al. (2011), analogous to the Procrustes method of superposition used in Fink, et al. (2005) and Schaefer et al. (2005). First, a group-wise registration strategy was used to generate a minimally biased nonlinear average of the entire group. Briefly, all faces were first normalized to the average linear dimensions of the group under study through the exhaustive estimation of all pairwise 12-parameter (3 translations, rotations, scales, and shears). Once each subject was transformed to the average dimensions of the group, a first voxel-by-voxel group average was estimated. Each subject was then nonlinearly registered to this first average and then a new voxel-wise average was generated. This procedure was then continued such that a higher resolution transformation was estimated at each nonlinear stage. Thus, each transformation maps the craniofacial features of each individual to the average craniofacial features of the group, (see average face in Figure 1). Second, 56 landmarks were placed on this population average at anatomically defined locations of the face (Figure 2). Third, the landmarks were warped back to each subject’s face using the inverse of the nonlinear transformation described above. This step provided a set of landmarks (and relevant distances) for each subject’s

face. Accuracy of projecting (or warping) landmarks from the average face onto each individual's face (using the non-linear registration of the faces) was evaluated by calculating Euclidean distances between the position of a projected landmark and the mean position of the same landmark placed manually on the same face (Chakravarty et al. 2011). While these distances varied between 1.58 and 10.80 mm, their standard deviation varied only between 0.62 and 2.61 mm. We also calculated the coefficient of variation (SD/Mean) for each of the 17 landmarks placed by the same observer on 10 different faces; these coefficients varied between 0.09 and 0.61. Note that this high intra-observer variability in placing landmarks on each individual face manually is eliminated by projecting the landmarks on each face automatically using the non-linear registration procedure.

Using the nonlinearly warped landmarks, the face of each individual was simulated by estimating a thin-plate-spline warp from the set of landmarks defined on the average face to the landmarks that were transformed using the nonlinear transformation that matches each subject to the 'average' face. This transformation was applied to a surface-based representation of the 'average' face in order to create a final image that closely resembled the face of the study participant. This final face image concealed the identity of the SYS individual but preserved the configuration of facial features unique to that individual at the resolution corresponding to the nonlinear registration (see an example image in Figure 3).

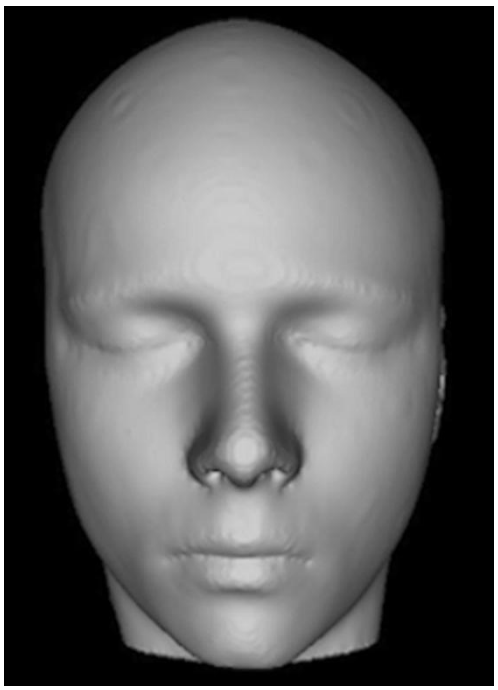


Figure 1:. An 'average' face was created using 597 MR-derived images from the Saguenay Youth Study (reprinted with permission from Chakravarty et al., 2011).

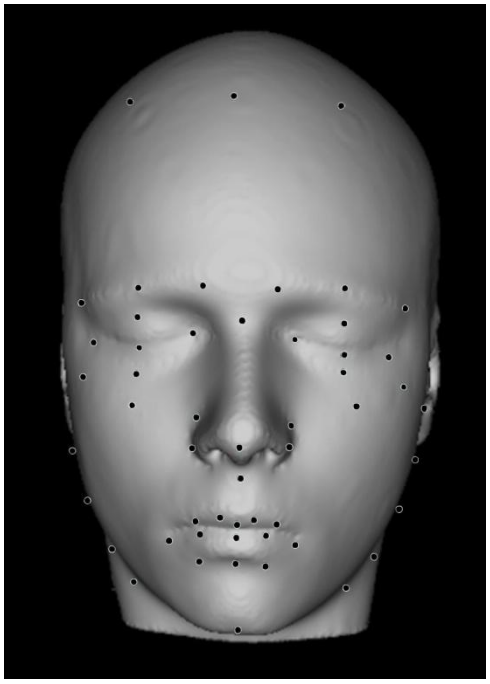


Figure 2: A total of 56 anatomical landmarks and semi-landmarks were defined using the average face. These were then warped back by non-linear registration to each individual SYS adolescent face, and the x, y, z coordinates were determined (reprinted with permission from Chakravarty et al., 2011).

Figure 3 A representative example of a face stimulus presented to raters for identification of its sex.

2.2.2.5 Quantification of facial features

Landmarks placed on the model image were warped back to individual faces using the inverse nonlinear transformation determined for each individual (Chakravarty et al., 2011). The x, y, and z coordinates, located in the standard Cartesian space, of all 56 landmarks in the 520 images were subjected to Principal Component Analysis (PCA) to investigate variation in facial features (Chakravarty et al., 2011). Each subject's scores for each principal component (PC) was then determined.

2.2.3 Study materials

In order to present a manageable number of faces to each rater, we created two sets of 270 MRI-reconstructed faces. Each set included 260 individualized faces and 10 copies of the average face randomly interspersed among the other faces. Both sets of face images depicted

faces of adolescents whose ages ranged from 12 to 18 years, and were composed of an equal number of male and female faces. The two image sets did not differ in terms of mean age of face presented, with equal age distributions across males and females. Set 1 was comprised of 126 males, (mean age 15.1 years, SD = 0.42) and 134 females (mean age 15.4 years, SD = 2.05 years). The second set included 129 males (mean age 15.4 years, SD = 1.89 years) and 131 females (mean age 14.95 years, SD = 1.82 years). The two sets of 270 images were uploaded to an online website (www.surveymonkey.com) that provides a user-friendly tool for creating online questionnaires. The order of images was randomized within each set.

2.2.4 Procedure

Upon recruitment, raters were contacted by e-mail with instructions that directed them to the online questionnaire. They were told that this was a study investigating perception of faces and sex of faces. Raters completed the experiment on their own computers, at a location of their choice and at a time of their convenience. At each trial, raters were presented with a single face (480 x 480 pixels), and asked to select the sex identity of the face in a forced-choice format. Raters had the option of selecting “male” or “female” for each face. Each image (270 in total) was presented on a separate screen. Only once rated, a subsequent face appeared.

2.2.5 Analysis

All statistical analyses were performed using JMP IN 8.0 (SAS Institute Inc., Cary, NC). Effect sizes are reported as Cohen's *d* (Cohen, 1988).

2.3 RESULTS

The following results are based on ratings from a sample of 60 White Caucasian females whose mean age was 18.7 years, SD = 1.21 (see Appendix 4 for the effect of rater's sex on identification accuracy).

2.3.1 Identifying the sex of individual faces

A greater proportion of face stimuli were rated as being 'male' (60% of faces) than 'female' (40% of faces), χ^2 (1) = 598.03, $p < 0.001$, although equal numbers of male and female faces were presented to raters. Male faces were correctly identified more frequently than female faces, as male faces received a greater proportion of correct responses (M = 70.1% correct responses, SD = 23.65) than female faces (M = 50.8% correct responses, SD = 25.27; t (516) = 9.45, $p < 0.001$, *d* = 0.79; Figure 4).

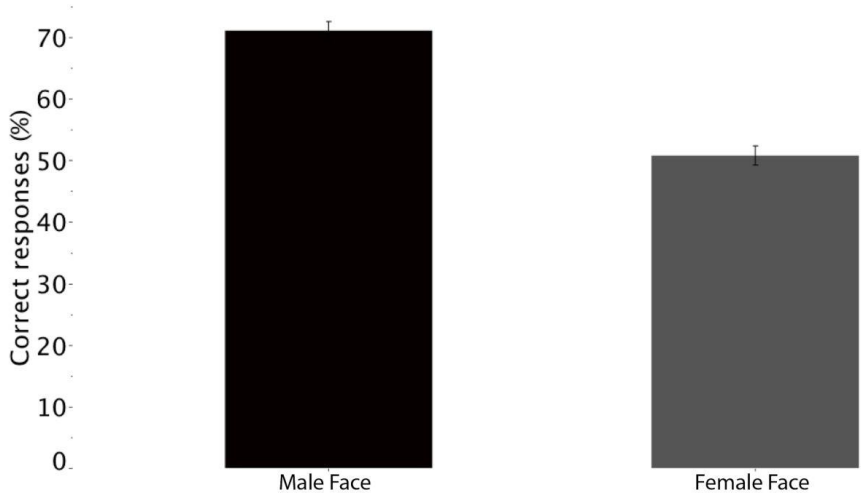


Figure 4:. Male faces were identified more correctly relative to female faces (t (516) = 9.45, p <0.001). Data presented are mean proportions of correct responses received by each sex group ± 1 standard error.

In addition to the main effect of sex on the number of correct responses (F(1, 515)= 99.6, p <0.001), we also observed a significant main effect of age (F(1, 515) = 32.6, p < 0.001). The sex of older faces was identified more accurately than that of younger faces. Sex of the face interacted with age (F(1, 515) = 18.68, p <0.001), reflecting an increasing distinctiveness of male faces with age. Whereas the sex of faces in the 12 year-old cohort was identified at chance levels (Figure 5), male faces 17 to 18 years old were more easily identified. Female faces appeared to be identified at chance levels across all ages. This suggests that perception of sex identity relies strongly on the presence or absence of male-related facial cues, which become more distinct with age.

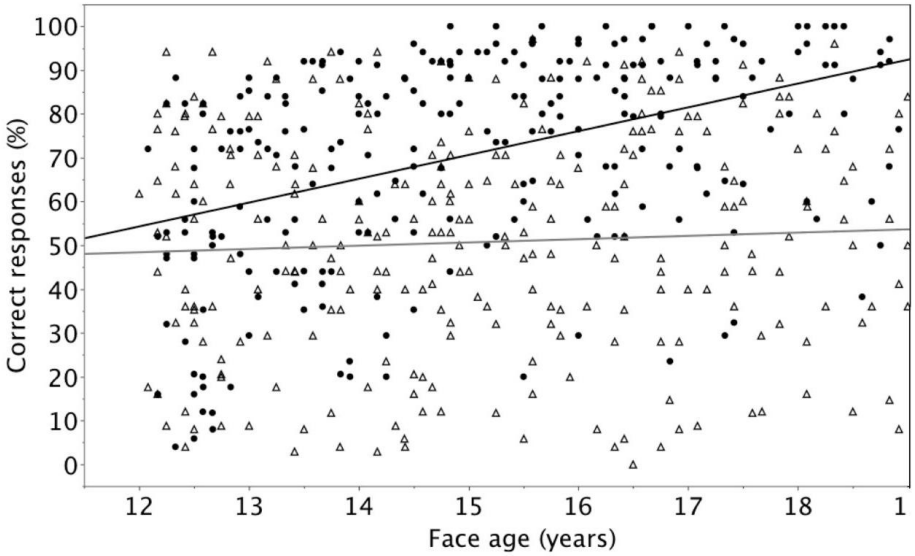


Figure 5:. The sex of male faces (filled-in circles) was identified more correctly with age ($r = 0.43$, $p < 0.001$). The sex of female faces (hollow triangles) was identified at chance levels across all age cohorts ($r = 0.058$, $p = 0.35$).

2.3.2 Relationship between ratings and facial features

In an effort to determine which facial features facilitated correct perception of sex, we examined the relationships between the principal components derived by Chakravarty et al. (2011) and the raters' accuracy scores in the current study. As we determined previously, five principal components (PCs) accounted for 70% of facial variation and were strongly related to age, sex and their interaction (Chakravarty et al., 2011). Simulations of these principal components are illustrated in Figure 6. Description of these PCs and their relationship with the identification accuracy are provided in Table 1.

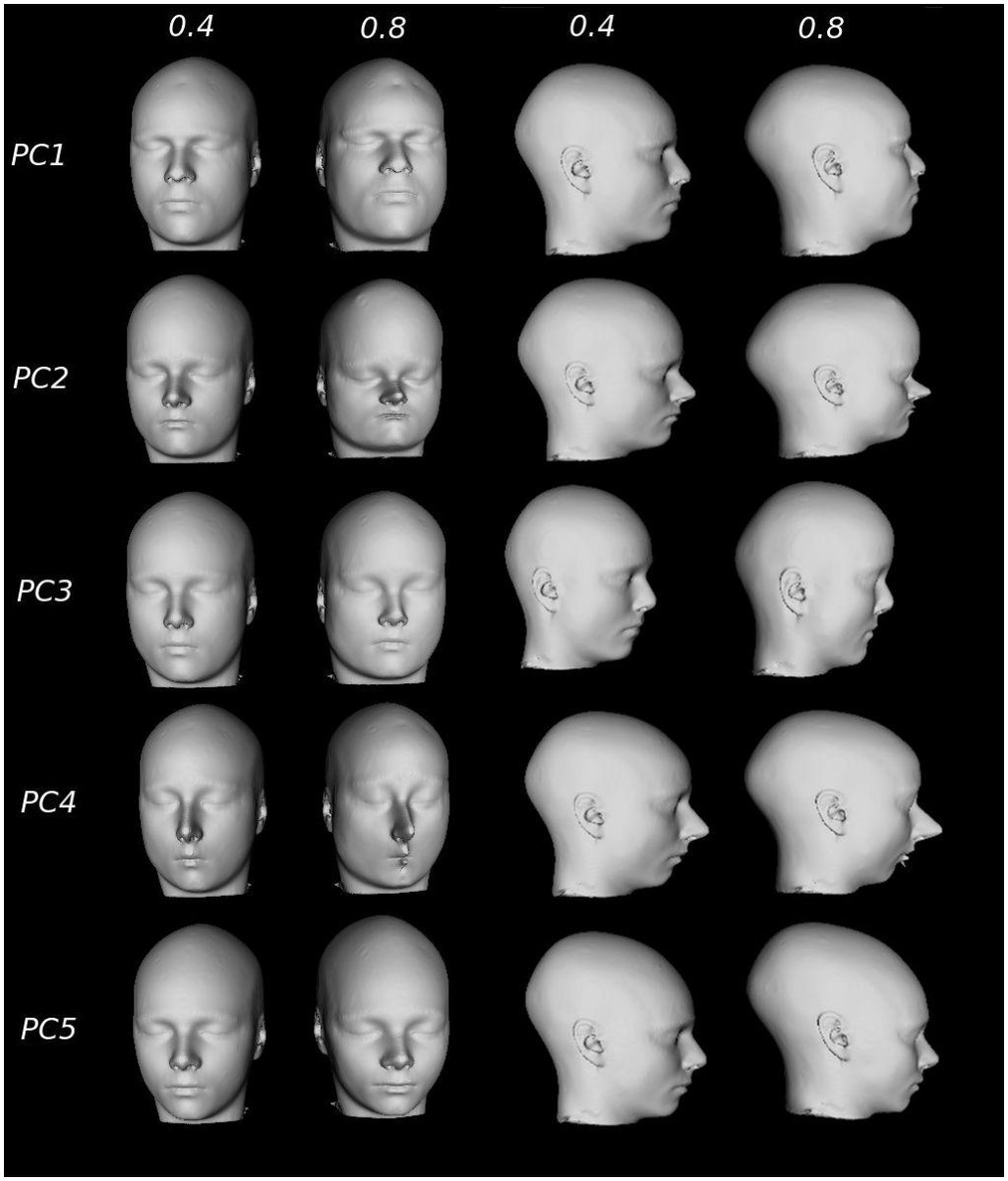


Figure 6: The average face was warped using landmarks defined (see Fig. 1) to create a facial feature simulation by warping each of PC 1-5 using 0.2, 0.4, 0.6, 0.8, and 1 of PC scores (reprinted with permission from Chakravarty et al., 2011).

Table 1: Effect of PC score on identification accuracy of male and female faces.

PC	%	Description	Identification Accuracy	
			Male faces	Female faces
1	38.5	broadening of the forehead, chin, jaw, and nose	**r = 0.50, p < 0.001**	**r = -0.62, p <0.001**
2	14.6	increased prominence of the forehead and decreased distance between facial features	**r = -0.18, p = 0.005**	**r = -0.17, p = 0.007**
3	8	enlargement of the brow line, broadening of the zygomatic arch and a more prominent jaw and chin	r = -0.07, p = 0.206	r = -0.05, p = 0.372
4	4.7	broadening of the chin, narrowing of the jaw and mouth, elongation of the nose, and a retreating jawline	**r = -0.34, p < 0.001**	r = 0.10, p = 0.098
5	3.5	narrower cheekbones, fuller but narrower lips and a less prominent jawline	**r = -0.14, p = 0.022**	**r = 0.19, p = 0.003**

Percentage scores in the second column describe the portion of variance in facial features the particular PC explained. Values presented are Pearson's r., significant correlations between identification accuracy and PC score are in bold.

High PC1 loading scores, characterized by broadening of the forehead, chin, jaw, and nose, strongly correlated with identification accuracy of male faces ($r = 0.50$, $p < 0.001$; Figure 7), whereas low PC1 loading scores most strongly correlated with identification accuracy of female faces ($r = -0.62$, $p < 0.001$). Low PC4-loading scores, described by narrower chin, shorter nose, broader jaw and mouth, and a more pronounced jawline, strongly correlated with identification accuracy in males (r = -

0.34, $p < 0.001$) but not in females. While PC2 and PC5 features also contributed somewhat to the correct identification of the face sex, this was not the case for PC3-based features.

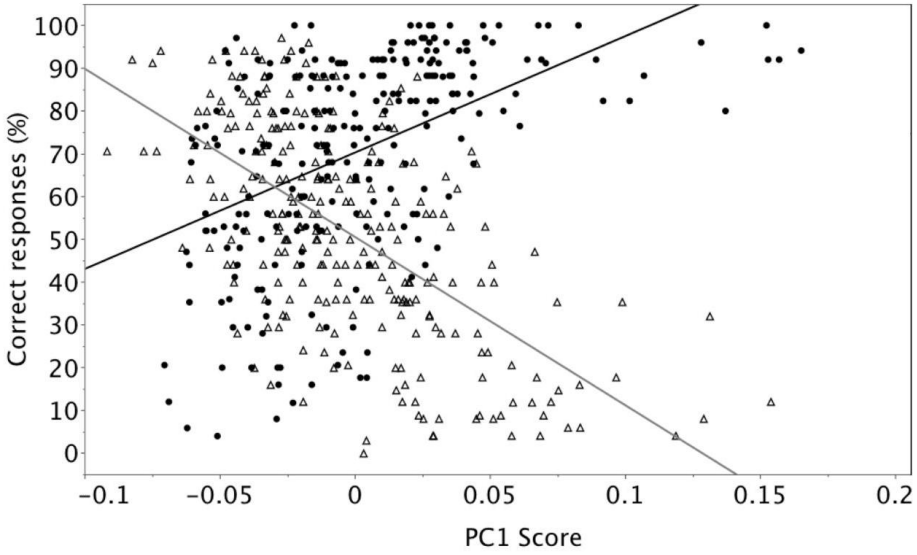


Figure 7: Higher PC1 scores in male faces facilitated identification accuracy by raters (filled circles, black line; $r = 0.50$, $p < 0.001$). Conversely, low PC1 scores in female faces facilitated identification accuracy (hollow triangles, grey line; $r = -0.62$, $p < 0.001$).

2.3.3 Does testosterone mediate the development of male-like features in males?

The raters' increased ability to identify the sex of older, relative to younger, male faces suggests that the older faces possessed more informative and reliable perceptual cues of sex. It is possible that the increase in perceived masculinity of facial features in males reflects the effects of puberty-related increases in testosterone levels. Plasma testosterone levels in pre-pubertal boys start to increase from the age of 11 years and continue to rise throughout puberty (Boyar et al., 1974; August et al., 1972). The mean plasma levels of bioavailable

testosterone were 9.02 ± 5.65 nmol/L in males, and 0.52 ± 0.36 nmol/L in females. The correlation between testosterone levels and age was significant in males ($r=0.66$, $p < 0.001$) but not in females ($r = -0.02$, $p = 0.73$).

To investigate whether testosterone levels influenced perceived masculinity of the male adolescents' faces, we determined whether bioavailable testosterone obtained in a given adolescent predicted correct identification of his face. Indeed, faces of males with higher bioavailable testosterone levels were more likely to be identified correctly as male ($r = 0.41$ $p < 0.001$). The effect of testosterone on identification accuracy was present even after controlling for chronological age ($r = 0.22$, $p = 0.0013$; Figure 8). A multiple regression analysis confirmed these independent effects of age and testosterone on identification accuracy and showed that testosterone was a stronger predictor of identification accuracy ($ß = 0.29$, $p < 0.001$) than age ($ß = 0.18$, $p = 0.038$). Testosterone and age together explained 18% of the identification accuracy variance (Adj $R^2 = 0.18$, $F(2, 201) = 22.93$, $p < 0.001$).

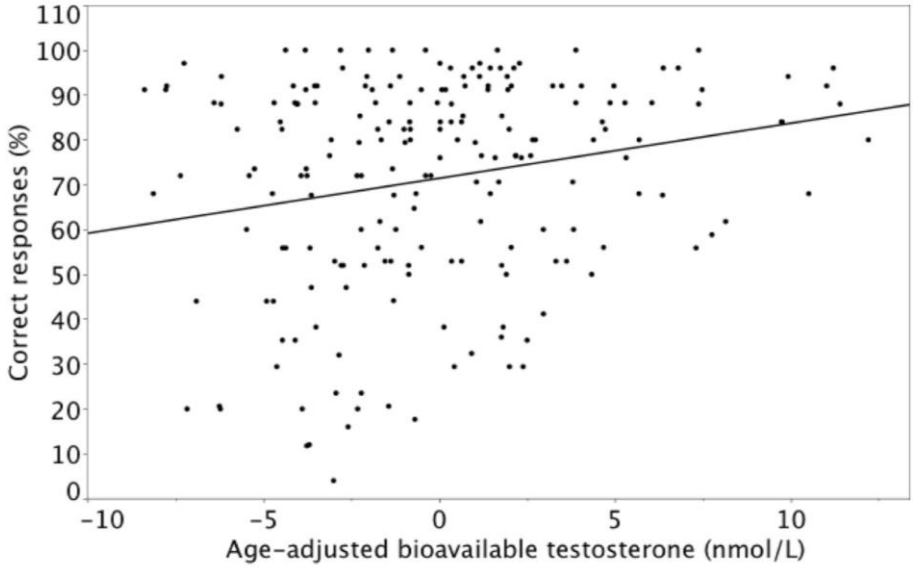


Figure 8: Males with higher levels of (age-adjusted) bioavailable testosterone were identified with greater accuracy by raters (r = 0.22, p = 0.001).

The effects of testosterone on facial development may be mediated by *AR*. To determine whether males with fewer *AR* CAG repeats were more likely to be perceived as 'male', we categorized males into those possessing more CAG repeats than the median number of CAG repeats across the entire sample ("Short *AR*" males); males with less than or equal to the median number of CAG repeats were classified as "Long *AR*" males (see Methods). The mean plasma levels of bioavailable testosterone were 8.79 ± 5.61 nmol/L and 9.3 ± 5.69 nmol/L in males with short and long *AR* gene, respectively. A model testing the main effects of *AR* genotype and age-adjusted bioavailable testosterone, and their interaction, on the identification accuracy suggested that identification accuracy is associated with the level of age-adjusted bioavailable testosterone ($F(3,200) = 10.79$, $p = 0.001$) but not the *AR* genotype ($F(3,200) = 0.18$, $p = 0.67$) or the

interaction between bioavailable testosterone and *AR* genotype ($F(3,200) = 0.18$, $p = 0.67$).

Next, we evaluated the relationships between age-adjusted testosterone levels, *AR* and the five principal components best characterizing face morphology (with age as a co-variate). To do so, we conducted a multiple regression analysis for the top five PCs (see Figure 6). The results of the regression revealed two strong predictors that could (together) explain 12% of the variance in age-adjusted testosterone levels (Adj $R^2 = 0.12$, $F(5, 200)=6.33$, $p < 0.001$): PC4 was identified as the best predictor of age-adjusted testosterone levels (ß = -0.29, $p < 0.001$), followed by PC1 (ß = 0.15, $p = 0.027$). The remaining PC2, PC3 and PC5 did not show any significant association with age-adjusted testosterone levels (PC2: ß = -0.1, $p = 0.14$; PC3: ß = -0.08, $p = 0.23$; PC5: ß = 0.1, $p = 0.16$). We then asked whether the relationship between age-adjusted testosterone levels and PC4, a principal component that was most strongly associated with testosterone levels, was moderated by *AR* genotype. This analysis again indicated significant effect of age-adjusted testosterone ($F(3,199) = 19.58$, $p < 0.001$) but no effect of *AR* genotype ($F(3,199) = 0.59$, $p = 0.443$) or interaction between the two ($F(3,199) = 0.14$, $p = 0.706$). A correlation between age-adjusted testosterone and PC4 was significant in both short *AR* males ($r = -0.27$, $p = 0.005$) and long *AR* males ($r = -0.34$, $p = 0.001$).

Finally, the *AR* genotype groups appeared to differ in the relationship between PC4 loading scores and identification accuracy (Table 2): males with the short *AR* gene exhibited nominally stronger negative correlation between PC4 and identification accuracy ($r = -0.42$, $p < 0.001$) than males with the long *AR* gene ($r = -0.25$, $p = 0.008$). But the interaction between PC4 and *AR* genotype was not significant $F(3,226) = 1.94$, $p=0.16$).

Table 2: Effect of PC score on sex identification accuracy of males possessing the short or long *AR* gene.

PC	Short-AR Males	Long-AR Males
1	**0.495 (*p* < 0.001)**	**0.502 (*p* < 0.001)**
2	-0.161 (*p* = 0.074)	-0.186 (*p* = 0.051)
3	0.015 (*p* = 0.866)	-0.162 (*p* = 0.090)
4	**-0.418 (*p* < 0.001)**	**-0.249 (*p* = 0.008)**
5	-0.119 (*p* = 0.188)	-0.183 (*p* = 0.055)

Values are Pearson's *r*. Bolded values indicate significant relationships. PC1 and PC4 best predicted identification accuracy in males with either the short or long *AR* gene.

2.3.4 Why were some females misclassified?

Raters performed at approximately chance levels (50% correct) when asked to identify the sex of female faces. It seems that raters relied on three sets of features to identify female sex, as indicated by the correlations between the number of correct responses and the weights of specific PCs (Table 1): low loading on PC1 ($r = -0.62$, $p <$

0.001), low loading on PC2 (r = -0.17, p = 0.007), and high loading on PC5 (r = 0.19, p = 0.003).

To investigate the facial features that were most likely to be judged as female and those that were most likely to be judged as male, we compared female faces that were identified most consistently as 'female' with those that were perceived most consistently as 'male'. To do this, we grouped female faces that were correctly identified as 'female' by at least 80% of raters ("perceived-as-female" group), and those that were identified as male by at least 80% of raters ("perceived-as-male" group). This classification resulted in 22 females that were consistently perceived as females, and 23 females consistently perceived as males.

While these females did not differ in age (t (43) = -1.04, p = 0.3) or Tanner stage at time of assessment (X^2 (2) = 1.15, p = 0.56), they did differ in their loadings on PC1 and PC5. As predicted from the overall correlations between PC loading and identification accuracy (Table 1), female faces consistently perceived as those of males had higher ("male") PC1 loading scores (t (43) = 6.21, p < 0.001) and lower ("female") PC5 scores (t (43) = 2.63, p = 0.01) than female faces consistently perceived as female.

Females perceived as males had higher body-mass index (BMI) than females perceived as females (t (43) = 4.05, p = 0.002) and exhibited greater fat mass, as assessed with bioimpedance (t (43) = 3.26, p = 0.002). MR images are based on soft-tissue contrast, and

thus, body fat may possibly play a role in how faces are perceived. Females perceived as males demonstrated a positive relationship between fat mass and loadings of PC1 ($r = 0.7311$, $p < 0.001$). Such a relationship between body fat and PC1 was not present in females perceived as females ($r = 0.03$, $p = 0.9$).

It is possible that the increased fat mass in females-perceived-as-males drove higher PC1 loading scores and thus increased presence of male-like features. In the entire sample, we controlled PC loading scores for the effects of body fat by residualizing the effects of body fat on PCs. The residualized PC scores (see Table 3A) indicate that even after controlling for body fat, low PC1 loading scores remain the strongest predictors of female sex identity ($r = -0.49$, $p < 0.001$).

Table 3.A: Effect of PC score, residualized for bioimpedance, on identification accuracy of male and female faces.

PC	Proportion Correct Responses (%)	
	Male faces	Female faces
1	0.361 ($p < 0.001$)	-0.487 ($p < 0.001$)
2	-0.159 ($p = 0.014$)	-0.130 ($p = 0.039$)
3	-0.142 ($p = 0.027$)	-0.055 ($p = 0.384$)
4	-0.316 ($p < 0.001$)	0.063 ($p = 0.322$)
5	-0.039 ($p = 0.550$)	0.083 ($p = 0.192$)

Values are Pearson's *r*. PC1 scores best predicted high identification accuracy in males and females. In males, PC4 serves as a second strong predictor of face sex.

Table 3.B: Effect of PC score, residualized for bioimpedance, on identification accuracy of males with short or long AR.

PC	Short-AR Males	Long-AR Males
1	0.338 (p = 0.002)	0.382 (p < 0.001)
2	-0.149 (p = 0.115)	-0.173 (p = 0.081)
3	-0.048 (p = 0.615)	-0.234 (p = 0.017)
4	-0.401 (p < 0.001)	-0.249 (p = 0.011)
5	-0.035 (p = 0.708)	-0.051 (p = 0.609)

Values are Pearson's *r*. Controlling PC loading scores for bioimpedance alters PC1 effects on identification accuracy, but does not significantly influence the effects of PC4 on identification accuracy.

In a preliminary analysis of whether sex hormones in adolescent females influenced the identification of their faces as female by raters, we used the levels of bioavailable testosterone sampled at the time of the adolescents' assessment. An analysis of the whole sample of female faces showed that faces of female adolescents were more likely to be identified as those of males if they possessed higher bioavailable testosterone ($r = -0.25$, $p < 0.001$); this was the case also after adjusting testosterone levels for chronological age ($r = -0.25$, $p < 0.001$). Consistent with this finding and the "maleness" of the face being captured by PC1, the PC1 loading scores of the female adolescents were positively related to the age-adjusted levels of bioavailable testosterone ($r = 0.29$, $p < 0.001$; Figure 9).

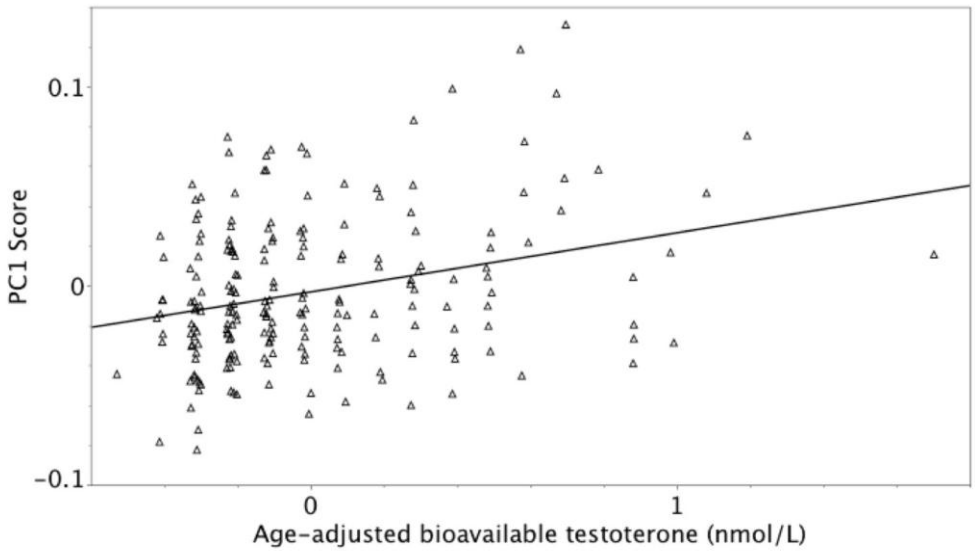


Figure 9:. Females whose faces exhibited high PC1 scores also exhibited higher levels of (age-adjusted) bioavailable testosterone ($r = 0.29$, $p < 0.001$).

2.4 DISCUSSION

In the present study, we used MRI-reconstructed faces developed by our group (Chakravarty et al., 2011) to identify the facial features that most strongly predict sex by a series of raters. One of the main advantages of using MR images is the preservation of the 3-dimensional representation of the face. We used this objective information in conjunction with subjective judgments to determine which objective features underlie perceptual cues about sex identity.

2.4.1 Facial features and perceived sex

Here we studied sexual dimorphism of the face and its development in a cross-sectional sample of healthy adolescents. Our results suggest that male and female faces appeared indistinguishable in terms of perceived sex until the late adolescence. This finding is consistent with results of Enlow (1982) who studied skeletal structure

and concluded that, before puberty, faces of boys and girls are very similar. Wild et al. (2000) also concluded that children's faces are less informative on the dimension of sex despite the fact that they appear as informative as those of adults vis-a-vis identity identification. Although Cheng et al. (2001) reported above-chance sex classification of children's faces, hey noted that sex classification of adult faces was more accurate than those of children. This slight discrepancy is most likely related to the fact that both our MRI-reconstruction based study and those based only on the skeletal structure lack information on the texture and colour of the skin, or shape of eye-brows; these facial characteristics provide additonal cues facilitating sex classification. The importance of these time-variant traits is illustrated, for example, by the fact that color of the skin contributes to male attractiveness more than morphological masculinity measures derived by 2D morphometric techniques (Scott et al. 2010). When using MRI-reconstructed faces, these additional cues are not present, leaving only local variations in the relative position, size and shape of the main craniofacial structures to guide the sex judgement.

Our results also indicate that raters' ability to classify sex of older faces was largely due to the raters' increased ability to identify older males. Raters did not identify older females as 'females' any better than they did younger females. This supports findings by Hoss et al. (2005) showing that masculinity in male faces but not femininity in female faces aids both children and adults in sex classification tasks.

Facial features described by PC1, namely increased breadth of the jaw, chin, nose, and forehead (Chakravarty et al., 2011) best predicted the sex of faces, in an age-dependent manner. The age-dependent change in the relationships between these features in males further suggests that, in males, age-related processes masculinized certain features. In females, this age-dependent change was not present. In males, features described by low loadings on PC4 and thus by wider jaw and mouth regions, a wider, more prominent jaw line, and shorter nose also predicted accurate male identification.

Previously, Bulygina et al. (2006) reported an enlargement of the face relative to the neurocranium during the first four post-natal years and no further changes in facial shape between 6 and 12 years of age. At this time, male and female individuals differed relatively little in terms of facial shape, and growth of their faces was relatively similar. Interestingly, while females attained maximal facial growth by the ages of 12-14 years, facial growth in males continued in 12 to 14 year old males. This resulted in the development of more pronounced brow ridges in males by young adulthood, and a larger, and more forward-protruding mandible (Bulygina et al., 2006).

It is possible that our raters' reduced ability to identify female faces was due to female facial characteristics signaling weaker female sex cues in absence of other sex-specifying cues (e.g., eyebrows, hair). While males take longer to achieve full facial development, once it is achieved, male features appear more distinct. Thus, older males

possessed features that more reliably cued ‘male’ identity. Bulygina et al. (2006) suggested that faces are sexually dimorphic as early as 6 months. A longitudinal study of Icelandic children found sex differences in several facial features as early as the age of 6 years, that is prior to puberty (Thordarson et al., 2006). These differences, however, do not seem to be perceptually apparent as our raters could not distinguish male and female faces at 12 years of age. More reliable perceptual cues, then, arise in adolescence, serving to distinguish masculine and feminine faces most clearly.

Our findings show a correspondence between objective sexually dimorphic changes in facial morphology and perception of sex identity in an age-dependent manner. This suggests that it is the maturational change that occurs at this time period that serves to define the sex of faces. Males seem to exhibit more distinct, robust facial changes with age relative to females, which correlate with more accurate identification of male faces with age.

2.4.2 What underlies sexually dimorphic facial changes in typically developing adolescents?

In this study, we have shown that males with higher levels of age-adjusted bioavailable testosterone were more likely to be perceived as male. Verdonck et al. (1999) showed that testosterone treatment (at low doses) triggers craniofacial growth in boys with delayed puberty. Testosterone effects on craniofacial structure were most pronounced in

the mandible, cranial base, and anterior face height (Verdonck et al., 1999).

Previously, Penton-Voak and Chen (2004) reported that men with higher levels of salivary testosterone appeared more masculine to raters. Pound et al. (2009) reported that testosterone levels measured 5 and 20 minutes after success in a competitive task were associated with masculine facial structure. Such a relationship between testosterone levels and perception of masculinity data could be supported also by data using our objective measures. This suggests that changes in testosterone levels may influence some aspects of facial shape and craniofacial development.

Testosterone effects on the development of the face may occur through changes in growth hormone (GH) levels. GH plays a major role in regulating growth, body development, and body composition. Craniofacial dimensions in children are also influenced by GH, and are altered in GH-deficient children (Spiegel et al., 1971). Growth hormone replacement can trigger the growth of mandibular ramus and of lower anterior facial height (Spiegel et al., 1971). Excess of growth hormone can result in gigantism during puberty or acromegaly in later life (Kashyap et al., 2011). Mandibular overgrowth, maxillary widening, and teeth separation are the most apparent characteristics of an acromegalic face (Kashyap et al. , 2011).

Previous studies have shown that GH secretion can be enhanced by testosterone treatment (Giustina et al., 1997; Loche et al., 1997;

Keenan et al., 1993). This relationship may support the changes we observed in the faces of males throughout adolescence. Plasma testosterone levels in pre-pubertal boys begin to increase at the age of 11 years, and continue to increase throughout puberty (Boyar et al., 1974; August et al., 1972). It is possible that the period of growth in males is especially punctuated by an increased presence of testosterone during puberty. This would result in apparent sex differences in the face. According to our study, it is these differences that can be identified by others.

2.4.3 Sex misclassification

Females that were most consistently perceived as males possessed significantly higher PC1 loading scores and significantly lower PC5 loading scores relative to females consistently perceived as female. Our analyses showed that high PC1 loading scores and low PC5 loading scores were strongly correlated with correct identification of male faces (Table 1). This finding suggests that misclassified females possessed more ‘male-like’ (PC1) and fewer “female-like” (PC5) features than other females in the group. This finding is consistent with the results obtained by O’Toole et al. (1998) who reported that sex-based classification of female faces with more masculine features was much slower than those of female faces with less masculine features.

Our current study also reports higher BMI, higher fat mass, and positive relationship between fat mass and loadings of PC1 in females

perceived as males. While our study is limited to analyses of facial shape and not craniofacial bone structure, a significant relationship between body fat and craniofacial morphology was also reported in an X-ray study by Sadeghianrizi et al. (2005). These authors showed that obese adolescents (male and female) had significantly larger mandibular and maxillary dimensions relative to healthy controls. In our study, wider mandibular dimensions were most strongly described by high PC1 loading scores, also associated with identification of male sex. A positive relationship between BMI and masculinity, dominance and low digit-ratio, was also reported in Schaefer et al. (2005).

In a preliminary analysis, we also showed that higher loadings of PC1 in females were related to higher levels of bioavailable testosterone (Figure 9). Increased testosterone might be related to increased body fat in misclassified girls. McCartney et al. (2007) showed that peripubertal obesity is associated with hyperandrogenemia and hyperinsulinemia throughout puberty. Rosenfield (2007) suggest that obesity in girls increases the likelihood of developing Polycystic Ovary Syndrome, a disorder associated with hyperandrogenism. Indeed, Baer et al. (2007) studied 8 to 10 year old females longitudinally for 7 years and found that higher BMI at childhood predicted higher levels of dihydroepiandrosterone sulphate in young adulthood. A follow-up study may examine whether other ovarian and/or adrenal androgens correlate with body fat and increased ‘male-like’ appearance among females.

Higher loadings of PC1 in female faces perceived as those of males provided evidence that testosterone influenced PC1-related features similarly in both males and females. Higher body fat among females perceived as males may have led to apparently larger PC1-features, leading to 'male' perceptions by raters. Alternatively, it is also possible that the relationship between body fat and presence of androgens in females perceived as males increases the development of more male-like craniofacial features and facial shape.

2.4.4 Limitations

The current results are based on ratings obtained in female raters only. Nonetheless, as explained in the Appendix 4, results obtained in the (smaller number) of male raters were not substantially different from those provided by female raters, except for the overall lower accuracy by the former. Furthermore, we have not considered the phase of menstrual cycle or the use of hormonal contraceptives by the female raters; both may influence face processing (e.g. Penton-Voak et al. 1999). Finally, recoding reaction times would complement the accuracy data and might provide additional insights not available in the present study.

2.5 CONCLUSIONS

This study used a novel method for studying perception of face sex and showed that reliable cues of male sex can be identified in faces, and that these become perceptually distinct with the progression

of puberty. This study used a unique set of face stimuli that stripped non-face sex-specifying cues to determine the aspects of face shape that most strongly indicate the sex of faces. In both males and females, perceived sex was influenced by levels of testosterone above and beyond chronological age. Levels of testosterone were also associated with changes in specific objectively-defined features. This link between the effects of testosterone on objective and subjective aspects of the face opens up the possibility of detecting clinically relevant phenotypes associated with abnormal levels of sex hormone.

The next chapter will study features of the skull and try to clarify to what extent it is the fat and to what extent it is the actual skull shape that contributes to the development of sex differences in the face.

CHAPTER 3

Does skull shape mediate the relationship between objective features and subjective impressions about the face?

3.1 INTRODUCTION

Sex identification plays an important role in social cognition (Macrae et al., 2002). Meeting a male vs. a female often triggers different expectations and social interactions. For example, sexually dimorphic features of the face contribute to the estimation of an individual's value as a mate (Thornhill & Gangestad, 1999). It has been suggested that cosmetic surgeons enhance attractiveness by creating a "hyper feminine" face and that a more squared-off jaw makes a woman look both more masculine and older (Adamson & Zavod, 2006). What are the sexually dimorphic features that make us perceive a face as a male or a female? Are these features determined by the skull? This study explores the role of skull shape in subjective impressions about the sex of a face.

Sexual dimorphisms in craniofacial morphology have been described in both skulls (e.g. cephalometry: Weston et al., 2007; X-rays: Bulygina et al, 2006) and face surface (e.g. photographs: Carre and McCormick, 2008, Penton-Voak et al., 2001; laser scanning: Toma et

al., 2008). Our previous studies used faces derived from magnetic resonance images (MRI) to explore the development of sex differences in the face during adolescence (Chakravarty et al., 2011; Mareckova et al., 2011). We used a surface-based warping technique to simulate the craniofacial anatomy of individual subjects by using a procedure that warped an average surface representation of the group under the study (Chakravarty et al, 2011). Using this approach, we showed that levels of age-adjusted bioavailable testosterone predicted face shape of both males and females (Mareckova et al., 2011).

Objectively measured features of the face and their development during adolescence are accompanied by variations in subjective impressions about the sex of these faces (Mareckova et al., 2011). We showed that males with broad jaw, nose, forehead, and narrow eyes were correctly classified as males, and females lacking these features were correctly classified as females. The total amount of body fat, however, correlated with the presence of these male-like features (Mareckova et al., 2011). Therefore, using the surface-based set of facial landmarks did not allow us to distinguish whether classifications of the individuals' sex were related to differences in craniofacial bone structure or differences in the soft tissue (fat and muscle) in the face (Mareckova et al., 2011).

The current study was aimed at determining the extent to which the skull or facial tissue influences impressions about the sex of a face. We placed landmarks on a number of skull elements visible on MR

images and hypothesized that skull features will mediate the relationship between facial features and subjective impressions about the sex of the face.

In addition to the faces included in our previous report (Mareckova, et al, 2011), we added another 401 faces, yielding a total of 876 faces. Bootstraping was used to perform mediation analysis and determine whether skull mediated the relationship between face shape and identification accuracy. Subsequently, we explored the role of body fat on the correct identification of the sex of a face and on the mediation by skull.

3.2 METHODS

3.2.1 Faces

Faces were derived from MR images obtained in a community-based sample of typically developing adolescents (age range = 12 to 18 years; M=180.18 months, SD=22.06; n=1,024) recruited in the context of the Saguenay Youth Study (SYS; Pausova et al, 2007). Faces were reconstructed successfully in 876 adolescents (411 males, 462 females; information about the sex of 3 participants was missing). Using this dataset, we derived information about facial features, skull features, and subjective impressions about the sex of these faces as described below.

To eliminate the influence of the brain in the subsequent image processing, we first extracted (and removed) the brains using an in-

house pipeline that requires nonlinear registration to a brain mask on a target (Avants et al., 2008). Then, a single participant was chosen to be an arbitrary reference and all other participants were rigidly aligned (3 translations and rotations) to that reference. MR images from all participants were corrected for intensity inhomogeneities (Sled et al., 1998) and clamped to an intensity range of 0-10,000. Next, the population was normalized to the average head size of the entire group. For each participant, this required matching to all other participants through the estimation of all possible 12-parameter transformations (3 shears, 3 scales, 3 translations, and 3 rotations). These transformations were then averaged and applied. Subsequently, a voxel-by-voxel intensity average was created to represent the average craniofacial structure of the group. Finally, a hierarchical iterative group-wise nonlinear registration procedure was initiated where each individual was nonlinearly registered to the population average, leading to the creation of a new average for the subsequent iteration. This process occurred in a hierarchical iterative fashion such that larger deformations were accounted for at earlier iterations and smaller deformations at every subsequent iteration. See Chakravarty et al. (2011) for more details regarding the image processing.

In order to capture variability in facial features among the 876 individuals, we placed 56 facial landmarks (Chakravarty et al., 2011) on the population-based average of 876 MRI-reconstructed faces and then warped these 56 landmarks back to each subject's face using the inverse of the nonlinear transformation; this allowed us to derive X, Y

and Z coordinates (and relevant between-landmark distances; Appendix 5) for all 56 landmarks for each of the 876 faces. These coordinates (located in Cartesian space) were mean-centered and subjected to principal component analysis (PCA) on covariances; principal components (PCs) explaining variability in facial features were generated for each individual.

3.2.2 Skulls

In order to capture variability in skull features among the 876 individuals, 19 landmarks were placed on the same population-based average at anatomically defined locations of the skull (Figure 1). While T1-weighted MR images provide very good contrast between soft tissues, bone tissue can only be estimated as the black non-tissue space. Placing skull landmarks in non-tissue spaces such as alveoli in the jaws enabled us to capture features that are independent of the amount of fat (or muscle) in the face. Landmarks were located in areas that enabled high precision in positioning (e.g. particular teeth, corners of the eye sockets, tip of the chin).

Next, the skull landmarks were warped back to each participant's head MRI using the inverse of the nonlinear transformation used to warp each individual into the population average. This step provided a set of landmarks (and relevant distances; Appendix 6) for each participant's skull. As above, the coordinates were mean-centered and subjected to PCA on covariances; principal components (PCs)

explaining variability in skull features were generated for each individual.

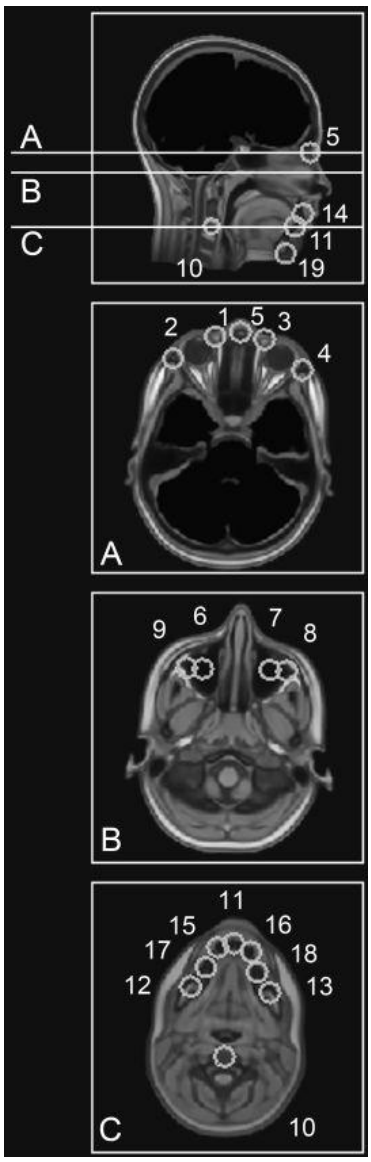


Figure 1: Skull landmarks (19) and their position:

1- Inside corner of the left eye socket;
2- Outside corner of the left eye socket;
3- Inside corner of the right eye socket;
4- Outside corner of the right socket;
5- Bridge of the nose;
6- Middle of the left mandibular sinus;
7- Middle of the right mandibular sinus;
8- Outside edge of the right mandibular sinus;
9- Outside edge of the left mandibular sinus;
10- Point around 2nd vertebrae where lower teeth most visible;
11- Front of the lower teeth;
12- Left jaw (unerupted lower 3rd molar);
13- Right jaw (unerupted lower 3rd molar);
14- Front of the upper teeth;
15- Left canine;
16- Right canine;
17- Left first molar;
18- Right first molar;
19- Tip of the chin bone.

3.2.3 Body fat

Saguenay Youth Study contains information about body fat, as measured with a multi-frequency bioimpedance (Xitron Technologies, San Diego, CA). This information was available for 831 out of the 876 participants with MRI-reconstructed face.

3.2.4 Raters

A total of 120 female first-year undergraduate students of Psychology from the University of Toronto (Toronto, ON, Canada) were recruited to rate the sex of MRI-reconstructed faces. Data from the first 60 female raters were already used in Mareckova at al (2011). Here we increased the sample size by another 60 raters and used the total of 120 for all further analyses. All raters were 17 to 30 years old (M=19.61, SD=2.44), of White (European descent) ethnicity (consistent with the ethnicity of the rated faces), and not taking antidepressant or antipsychotic medication at the time of their participation in the study. We chose to work with female raters only to avoid adding more variance in the ratings by including male raters; in a pilot study, we showed that male (vs. female) raters are less accurate in judging the face sex. Students participated for a course credit and a 1-in-6 chance to win a gift voucher for $50 CAD.

3.2.5 Rating of the face sex

In order to present a manageable number of faces to each rater, we divided the total number of faces into four batches. The first two batches of faces were rated in the context of Mareckova et al (2011) study and provided us with sex judgments about 475 faces included in the current model (n=876). Batch 3 and 4 were created later, specifically for this study, and included 195 faces each. The remaining 11 faces were excluded from the rating part of the study due to slight

distortions around the lips. The four batches did not differ in terms of sex of the faces (X^2=0.22, p=0.97) or age (F(3,861)=1.85, p=0.14) and we thus combined all four batches and analyzed all data in the context of both the facial features and skull features.

For consistency with Mareckova et al. (2011), 30 raters rated each face in batches 3 and 4 and the batches did not differ in terms of the mean age of face presented, with equal age distributions across males and females. Batch 3 included 92 male (mean age = 175.7 months, SD=20.13) and 103 female (mean age = 178.6 months, SD=21.8) faces. Batch 4 included 91 male (mean age = 175.7 months, SD=19.56) and 104 female (mean age = 178.8, SD = 21.96) faces. Images were uploaded to an online website (www.surveymonkey.com) that provided a user-friendly tool for creating online questionnaires.

Instructions and link for the online questionnaire were emailed to the recruited raters and they completed the experiment on their own computers, at time and location of their choice. Each trial (195 in total) appeared as a separate screen including a single face (480x480 pixels; see Figure 2 for an example face) and a forced-choice question about their impression about the sex of the face (male or female). There was no time limit per face.

Figure 2: An example of MRI-reconstructed face presented to the raters

3.2.6 Analysis

First, identification accuracy was calculated as the percentage of correct identifications of the sex of each face by the 30 raters. Next, we compared the newly derived principal components describing variability in the face morphology with principal components used in Mareckova et al (2011), and then used the newly derived principal components to study the relationship between objective facial features and identification accuracy (replicating the analysis performed in Mareckova et al., 2011).

Second, we used bootstrapping to perform mediation analyses. We hypothesized that skull features mediate the relationship between objective facial features and subjective impressions about the face. This mediation hypothesis was tested with a bootstrap procedure to determine the significance of the indirect effect (Preacher & Hayes, 2004). A total of 5,000 bootstrap resamples were used to provide stable estimates of the direct, indirect, and total effects. We determined 95%

confidence intervals from the bootstrap resamples and any interval that did not include 0 was considered to be significantly different from 0.

Subsequently, the possible role of body fat on these mediations was explored. We hypothesized that if we adjusted facial features for the total amount of body fat, skull would mediate the relationship between body fat-adjusted facial features and identification accuracy in both males and females.

3.3 RESULTS

3.3.1 Facial features

The 56 facial landmarks data were suitable for PCA (Kaiser-Meyer-Olkin (KMO) =0.932; Barlett's test of sphericity was significant (X^2=278744.160, df=14028, p<0.0001)). The PCA identified 10 principle components (PCs) that accounted for 76.6% of variance in the facial features. These 10 PCs are characterized in Table 1A.

Table 1A: Effect of face-based principal components (PCs) on identification accuracy of male and female faces (correlations that survived correction for 10 multiple comparisons are in bold).

Face PC	%	Description	Male faces	Female faces
1	41.2	Wide face, short nose, narrow eyes	**r= 0.41, p<0.001**	**r= -0.51, p<0.0001**
2	9.6	Narrow face with longer lower face	**r= 0.19, p<0.001**	r= -0.11, p=0.02
3	6.6	Wide cheekbones, short nose, longer lower face	r= 0.12, p=0.02	r= 0.02, p=0.74
4	4.3	Long nose, small mouth	**r= -0.30, p<0.001**	r= 0.11, p=0.02
5	3.4	Longer lower face, wide mouth	r= -0.01, p=0.84	r= -0.05, p=0.33
6	2.8	Big eyes, longer nose, small mouth	**r= -0.21, p<0.001**	**r= 0.35, p<0.001**
7	2.7	Small narrow face	**r= -0.14, p=0.004**	r= 0.01, p=0.77
8	2.2	Small eyes, short nose, longer lower face	r= 0.08, p=0.09	r= -0.03, p=0.51
9	2.0	Narrow face with small mouth and chin	**r= 0.17, p=0.0004**	r= -0.05, p=0.31
10	1.8	Long nose, small forehead	r= -0.11, p=0.03	r= -0.04, p=0.38

Percentage scores in the second column describe the portion of variance in facial features the particular principal component (PC) explained. Values presented are Pearson's r. Significant correlations between identification accuracy and PC score are in bold.

Table 1B: Effect of skull-based principal components (PCs) on identification accuracy of male and female faces (correlations that survived correction for 10 multiple comparisons are in bold).

Skull PC	%	Description	Male faces	Female faces
1	43.8	Wide cheekbones, short nose	**r= 0.42, p<0.0001**	**r= -0.49, p<0.0001**
2	12.4	Small jaw with less prominent chin	**r= 0.22, p<0.0001**	**r= 0.14, p=0.003**
3	8.9	Wide cheekbones, jaw; longer nose and lower face	**r= 0.26, p<0.0001**	**r= -0.18, p<0.0001**
4	6.1	Small sockets, disappearing chin	r= -0.05, p=0.28	r= 0.03, p=0.53
5	3.3	Wide cheekbones and prominent chin	r= -0.07, p=0.15	r= -0.08, p=0.10
6	2.8	Large socket distance, longer nose	r= -0.04, p=0.40	r= 0.11, p=0.03
7	2.5	Small face with big sockets	r= 0.02, p=0.64	r= 0.03, p=0.57
8	2.1	Narrow face with small features	r= -0.13, p=0.01	r**= 0.22, p<0.0001**
9	1.7	Small socket distance and lower face	r= 0.05, p=0.34	r= -0.005, p=0.91
10	1.4	Short nose, larger socket distance	r= 0.04, p=0.40	**r= 0.14, p=0.003**

Face PC1 reflected typically male features – wide face with short wide nose and narrow eyes (Figure 3A). Face PC2 was narrow with longer lower face and narrow eyes. Face PC3 was characterized by a wide face with big eyes and longer lower face. Face PC4 had typically female features - longer nose, small mouth and big eyes. Wide face, mouth, and bigger eyes identified face PC5. Face PC6 was also a female one - longer nose, small mouth, and big eyes - but was narrower than face PC4. Face PC7 was small and narrow, face PC8 had small features and eyes in particular, face PC9 was narrow, and face PC10 was characterized by a long nose but small forehead.

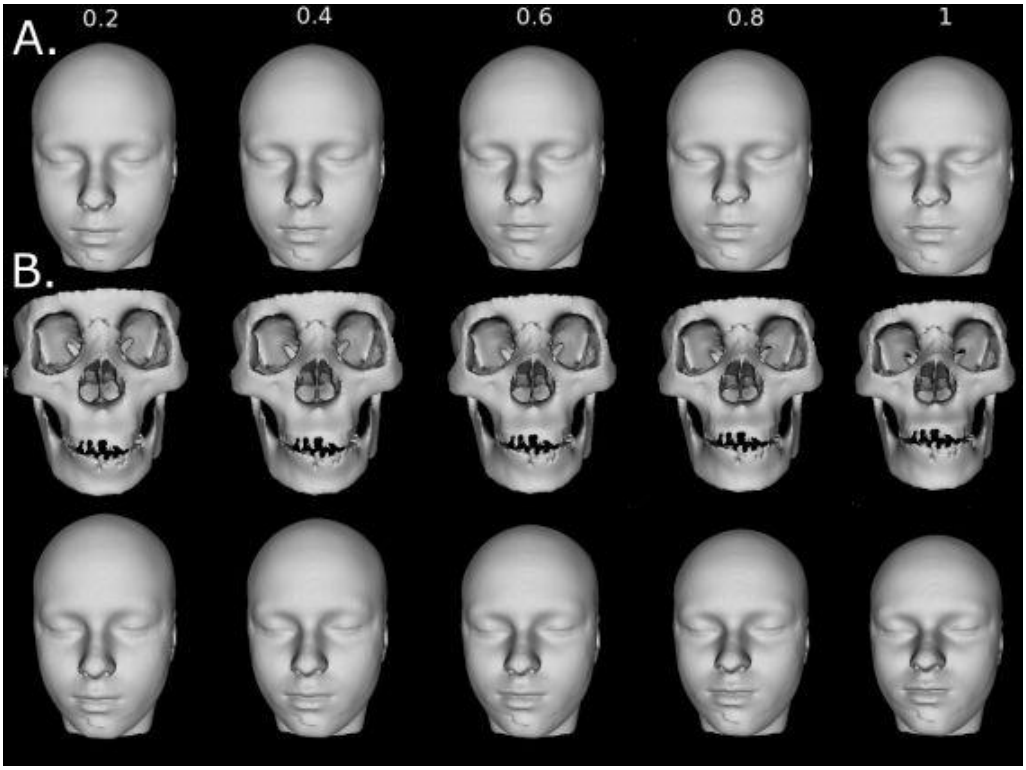


Figure 3A: Visualization of the face PC1: the average face was warped using facial landmarks to create a facial feature simulation by warping face PC1 using 0.2, 0.4, 0.6, 0.8, and 1 of face PC1 scores.

Figure 3B: Visualization of the skull PC1: the average skull was warped using skull landmarks to create a skull feature simulation by warping skull PC1 using 0.2, 0.4, 0.6, 0.8, and 1 of skull PC1 scores. The top row shows the PC1 features of the skull, the bottom row shows the skull PC1 features projected on a face.

3.3.1.1 Comparison of facial features derived from the original (Mareckova et al., 2011) and current (n=876) sample of faces

The current face PCs correlated well with the face PCs derived for the original subsample reported in Mareckova et al. (2011). An overview of these correlations is provided in Appendix 7. There were large correlations between the current and original PC1 ($r=0.86$, $p<0.0001$), current PC2 and original PC3 ($r=-0.86$, $p<0.0001$), current PC3 and original PC2 ($r=-0.82$, $p<0.0001$), and current and original PC4 ($r=0.67$, $p<0.0001$). We can conclude that the high correlations in the first four PCs that account for most of the variability in facial features confirm our previous findings (Mareckova et al., 2011) in a larger sample (n=876); note that “original” PC2 (PC3) and “current” PC3 (PC2) captured, respectively, very similar face features.

3.3.2 Skull features

The 19 skull landmark data were suitable for PCA (KMO=0.878; Barlett’s test of sphericity was significant ($X^2=64978.425$, $df=1596$, $p<0.0001$)). The PCA identified 10 PCs that accounted for 85% of variance in the skull features. These 10 PCs are characterized in Table 1B.

Wide cheekbones and shorter nose were the main skull features captured by PC1 (Figure 3B). Skull PC2 had small jaw with less prominent chin. Skull PC3 was characterized by wide cheekbones and

jaw, accompanied with longer lower face and nose. Skull PC4 had small eye sockets and less prominent chin. Wide cheekbones and prominent chin characterized skull PC5, longer nose with eye sockets further apart characterized skull PC6, and small face with big eye-sockets characterized skull PC7. Skull PC8 had typically female features and narrow face. Skull PC9 had smaller lower face and short distance between the eye sockets, and skull PC10 had wider eye-socket distances and shorter nose.

3.3.3 Relationship between skull and face features

Skull-based PCs and face-based PCs were highly correlated (Figure 4): face PC1 – skull PC1 ($r=0.94$, $p<0.0001$) where skull PC1 mirrored the wide and typically male-like face PC1; face PC2 – skull PC2 ($r=0.75$, $p<0.0001$), with both PCs reflecting a male-like but narrow face; and face PC3 – skull PC3 ($r=0.84$, $p<0.0001$) where both PCs had wide cheekbones and longer lower face.

3.3.4 Effect of sex and age on development of skull and face features

Effect of Sex, Age, and their interaction on (1) face PC1 and (2) skull PC1 features was explored using two separate two-way ANOVAs. An interaction between Sex and Age was present for both face PC1 (ANOVA: $F(3,869)=6.37$, $p=0.0003$; Age*Sex: $t=2.93$, $p=0.003$) and skull PC1 (ANOVA: $F(3,869)=9.41$, Age*Sex: $t=2.98$, $p=0.003$). There was a positive effect of age on face PC1 ($t(410)=4.32$, $p<0.0001$,

R^2=0.04) and skull PC1 (t(410)=5.14, p<0.0001, R^2=0.06) in males but no effect of age on face PC1 (t(461)=0.43, p=0.67) and skull PC1 (t(461)=1.32, p=0.19) in females.

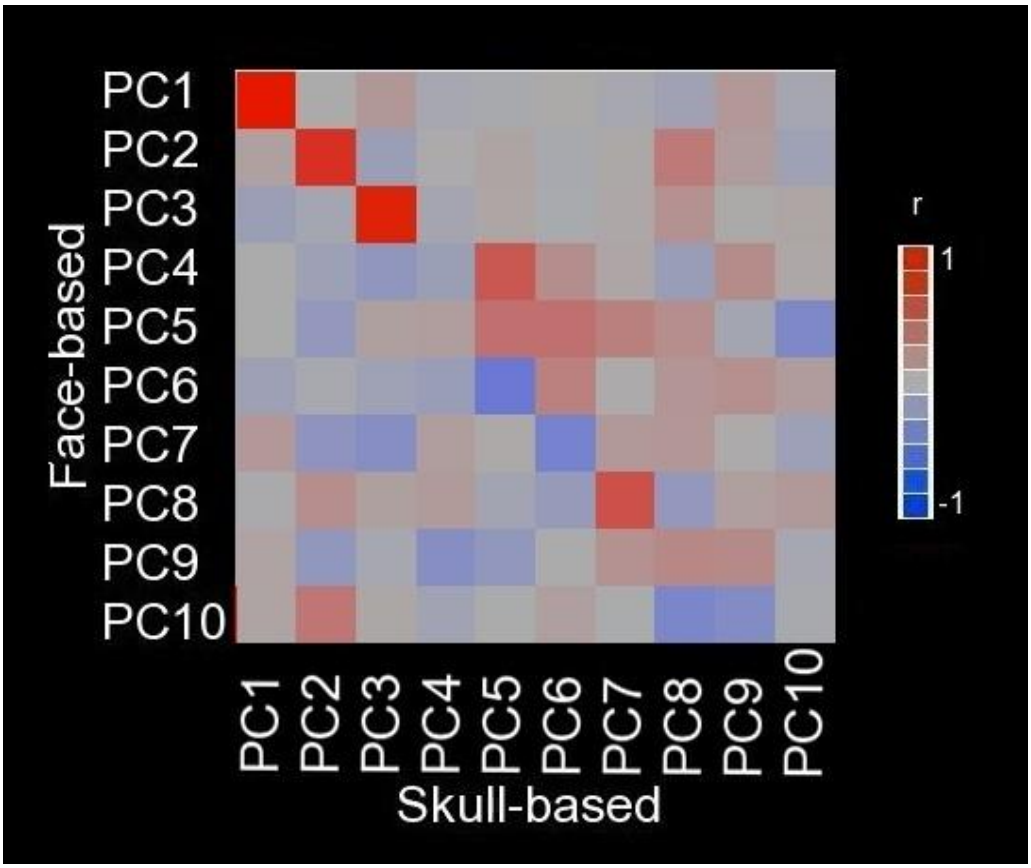


Figure 4: Correlation matrix of face and skull principle components (PCs)

3.3.5 Relationship between facial features and sex judgments about a face

Correlation analyses explored what features of the face are associated with correct subjective identification of the sex of a face. Our previous results showed that facial features captured by PC1 were the strongest predictor of correct identification of the sex of a face by raters (Mareckova et al., 2011). This relationship between facial features of PC1 and identification accuracy was also observed in the current larger

sample. The correlation between PC1 features (Females: M=-0.87, SD=19.3; Males: M=0.07, SD=19.48) and identification accuracy (Females: M=48.8, SD=24.84; Males: M=71.76, SD=23.13) was negative in females (r=-0.51, p<0.0001) and positive in males (r=0.41, p<0.0001).

3.3.6 Does skull mediate the relationship between facial features and sex judgments about the face?

To examine the possible mediation of the facial features – sex identification accuracy relationship by skull, bootstrapping meditation analyses were conducted separately for males and females.

In males (n=411), this analysis revealed that the relationship between face-based PC1 and identification accuracy was mediated by skull-based PC1, (ab=0.35, SE=0.16, 95% CI [0.03, 0.64]). Although the total effect of face-based PC1 on identification accuracy was significant, the direct effect of face-based PC1 on identification accuracy was not significant when the indirect path through skull-based PC1 was taken into account. Thus, in males, the relationship between face-based PC1 and identification accuracy was fully mediated by skull-based PC1 (Figure 5).

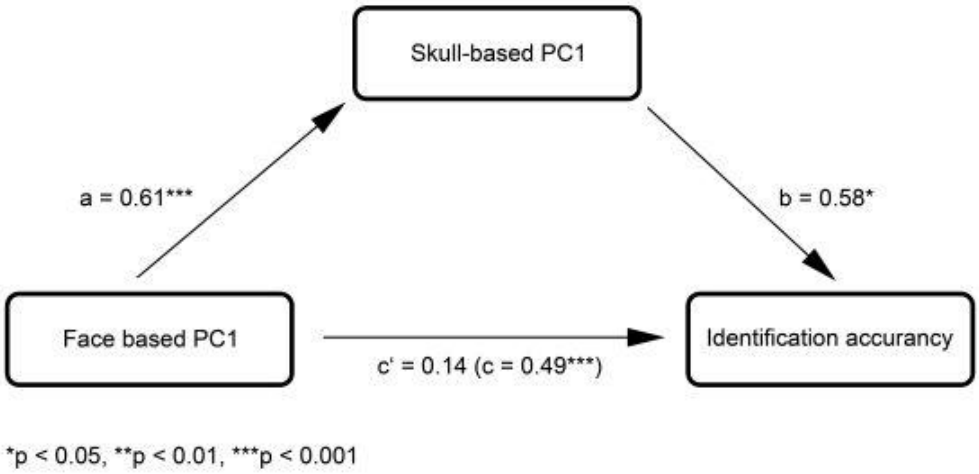


Figure 5: Skull mediates fully the relationship between facial features of PC1 and identification accuracy of male faces (ab=0.35, SE=0.16, 95% CI [0.03, 0.64]; a=0.61, SE=0.01, p=0.000; b=0.58, SE=0.25, p=0.02; c=0.49, SE=0.05, p=0.000; c'=0.14, SE=0.16, p=0.37; R^2=0.18, n=411)

In females (n=454), the relationship between face-based PC1 and identification accuracy was not mediated by skull-based PC1 (ab=-0.16, SE=0.14, 95 CI [-0.45, 0.12]).

Next, we explored whether the presence/absence of mediation described above might be related to the amount of body fat. As expected, female adolescents had more body fat than males (Females: n=445, M=24.54, SD=8.40; Males: n=386, M=14.91, SD=7.69; t(829)=17.14, p<0.0001, R^2=0.26). In each sex separately, we adjusted face-based PC1 by the total amount of body fat.

In males, the relationship between body fat-adjusted face-based PC1 and identification accuracy was mediated by skull-based PC1, ab=0.32, SE=0.10, 95% CI [0.14, 0.52]. The total effect of body fat-adjusted face-based PC1 on identification accuracy was reduced, when the indirect path through skull-based PC1 was taken into account. In

males, the effect of body fat-adjusted face-based PC1 on identification accuracy was thus partially mediated by skull-based PC1 (Figure 6).

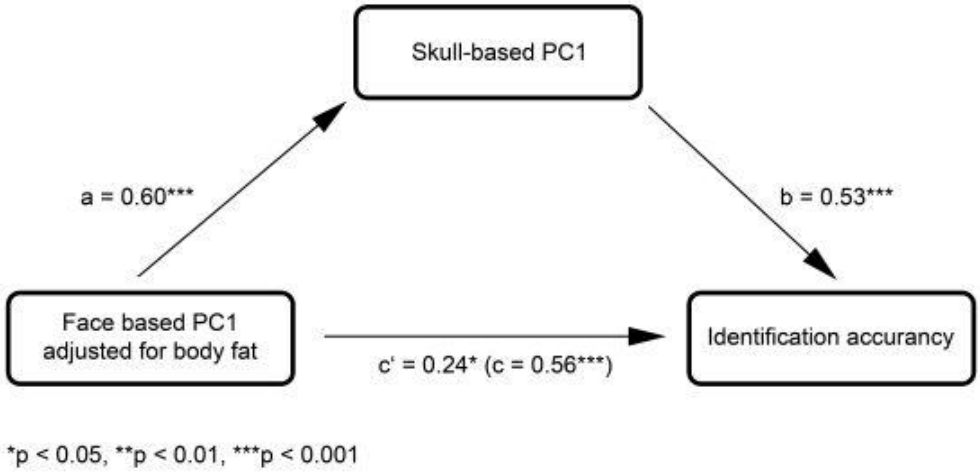


Figure 6: Skull mediates partially the relationship between body fat-adjusted facial features of PC1 and identification accuracy of male faces (ab=0.32, SE=0.08, 95% CI [0.14, 0.52]; a=0.6, SE=0.02, p=0.000; b=0.53, SE=0.16, p=0.000; c=0.56, SE=0.06, p=0.000; c'=0.24, SE=0.11, p=0.038; R^2=0.20; n=385).

In females, this analysis revealed that the relationship between body fat-adjusted face-based PC1 and identification accuracy was mediated by skull-based PC1, ab=-0.6, SE=0.07, 95% CI [-0.74, -0.47]. Although the total effect of body fat-adjusted face-based PC1 on identification accuracy was significant, the direct effect of body fat-adjusted face-based PC1 on identification accuracy was not significant when the indirect path through skull-based PC1 was taken into account. In females, the effect of body fat-adjusted face-based PC1 on identification accuracy was fully mediated by skull-based PC1 (Figure 7).

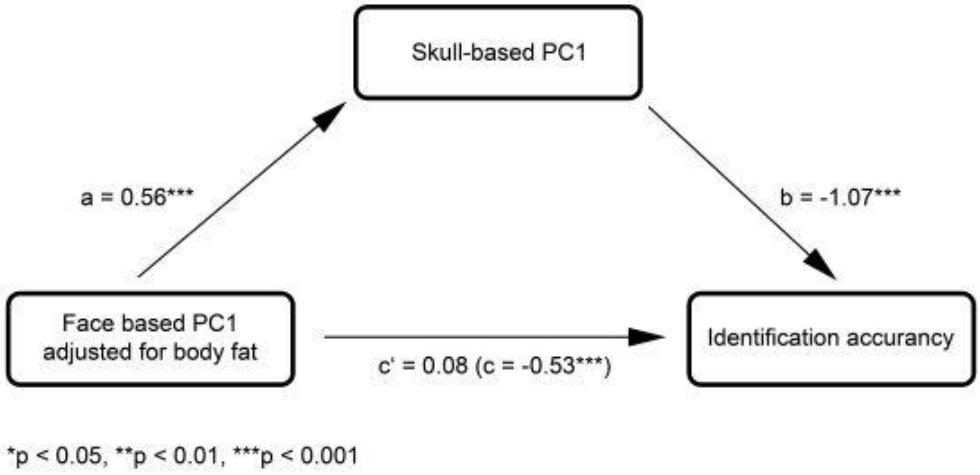


Figure 7: Skull mediates fully the relationship between body fat-adjusted facial features of PC1 and identification accuracy of female faces (ab=-0.6, SE=0.07, 95% CI [-0.74, -0.47]; a=0.56, SE=0.02; p=0.000; b=-1.07, SE=0.13, p=0.000; c=-0.53, SE=0.07; p=0.000; c'=0.076, SE=0.09, p=0.43; R^2=0.25; n=437).

3.3.7 The role of fat in recognizing sex of a face

Finally, we wanted to clarify whether the fat in the face is only noise that diminishes the sex-specific signal coming from the skull features and thus lowers the chances of correct sex recognition, or whether the fat might be actually a signal contributing to correct sex recognition.

Linear regression was used to explore the relationship between body fat, the estimator of fat in the face, and accuracy of sex judgments. Body fat contributed to the perceived maleness of both male and female faces; males with more body fat were more correctly classified as males (t(385)=2.02, p=0.04) while females with more body fat were more often *misclassified* as males (t(436)=-9.09, p<0.0001). These results suggest that body fat might facilitate correct recognition of a male face but could interfere with correct sex recognition of a female

face. The effect size was much larger in females compared with males. While body fat predicted 16% of variance in sex judgments about female faces, it predicted only 1% of variance in sex judgments about male faces. Having less body fat was thus essential for correct sex classification of a female face.

Face PC1 and skull PC1 were highly correlated in both males and females. Still, in males, skull features of PC1 ($t(385)=9.49$, $p<0.0001$, $R^2=0.19$) could explain slightly more variance in sex judgments than facial features of PC1 ($t(385)=9.1$, $p<0.0001$, $R^2=0.18$), and in females, facial features of PC1 ($t(436)=-12.41$, $p<0.0001$, $R^2=0.26$) could explain slightly more variance in sex judgments than skull features of PC1 ($t(436)=-11.91$, $p<0.0001$, $R^2=0.25$). It is thus no surprise that multiple regression exploring the role of skull, face, and body fat on accuracy of sex judgments (run separately in males and females) identified skull PC1 and, to a lesser extent body fat, as predictors of sex recognition in males ($F(3,382)=199.48$, $p<0.0001$; Skull PC1: $t=2.66$, $p=0.008$; Body fat: $t=-2.02$, $p=0.04$), but face PC1 and body fat as predictors of sex recognition in females ($F(3,433)=19.81$, $p<0.0001$; Face PC1: $t=-3.36$, $p=0.0009$; Body fat: $t=-3.63$, $p=0.0003$). In males, skull PC1 and body fat together could predict 19% of variance (Adj $R^2=0.19$). In females, the face PC1 and body fat together could predict 28% of variance (Adj $R^2=0.28$).

3.4 DISCUSSION

Using MR images, we studied the role of craniofacial bone structure in the relationship between objective facial features and subjective judgments about the sex of a face. We showed that skull features are essential for our perceptions about the sex of a male face. Skull mediated fully the relationship between facial features and sex judgments about the male face. This was not the case in females. Mediation by skull in females was revealed only after adjusting facial features for total body fat.

Fat contributed to correct sex recognition of a male face, but this effect was very small and, as we have shown in our mediation analyses, the skull itself could explain the presence of the relationship between facial features and sex judgments about male faces. In contrast, presence of fat reduced the correct sex recognition of a female face and this effect was large enough to prevent the skull features from mediating the relationship between facial features and sex judgments.

There seems to be two main cues that influenced raters' judgements of sex and suggested maleness of the face: (1) larger craniofacial bone structure and features of PC1 and (2) more body fat that may contribute to the facial features but mask, to some extent, the typically female features of the skull. We will now describe the sex differences in skull development and briefly explain the underlying biological mechanisms.

Sex differences in face shape emerge as early as in the first year of life and increase with age (children 6-11 years: 2.7% dimorphism; adolescents 12-17 years: 3.5% dimorphism; Bulygina et al., 2006). While the growth of female faces slows down by the age of 13 years and stops at the age of 15 years, male faces continue developing till late adolescence (Bulygina et al., 2006). Male faces are larger than those of females already before puberty but the development of typically male facial features continues till the age of 18 years (Bulygina et al., 2006). Prolonged growth of the male face and skull during puberty (Enlow et al., 1996) was observed also in our current study.

Typically male features of face PC1 were characterized by wider face, shorter nose, and narrow eye shape. Sex differences in the width to height ratio of the upper face (between the lip and the brow) also have been described by others: broader and shorter upper face in males (vs. females) was reported in both humans (Weston et al., 2007) and chimpanzees (Weston et al., 2004). During adolescence, male upper faces remain shorter than expected for their overall size (Weston et al., 2007) while lower face (mandible) becomes enlarged (Ferrario et al., 1998). Male, compared with female, faces were also characterized by greater bone strength and dimension, which is related to higher rates of periosteal bone formation (Vanderschueren et al., 2004). Presence of these larger skull features in males might possibly explain why our raters could recognize males more accurately than females (Mareckova et al., 2011), and why the skull mediated the relationship between facial features and sex recognition in males but not females.

Presence of male-specific features increases as a function of age-adjusted bioavailable testosterone (Mareckova et al., 2011). Aromatase converts testosterone to estrogen, the key hormone involved in skeletal growth and maintenance of bone mass (Riggs et al., 2002). Jaw area is rich in estrogen receptors (ERs), possibly explaining why ovariectomy (Tanaka et al., 1999) and estrogen deficiency (Ejiri et al., 2008) have a negative effect on jaw growth. Androgen deficiency in men and estrogen deficiency in postmenopausal women can induce bone loss (Vanderschueren et al., 2004). On the other hand, women with polycystic ovary syndrome (PCOS), which is associated with increased levels of androgens, have higher mineral bone density compared with age-matched controls (Buchanan et al., 1988; Dagogo-Jack et al., 1997). It is likely that not only fat but also higher levels of androgens influenced facial features of females that were *misclassified* as males in the sex judgment task of this current study.

Since (1) fat influences sex judgments about female faces and (2) meeting a male vs. a female often triggers different expectations and social interactions, are people approaching and treating misclassified females differently than correctly classified ones? Literature shows that it might be the case. Sexually dimorphic features of the face contribute to the estimation of an individual's value as a mate (Thornhill, & Gagenstad, 1999), and heterosexual males tend to prefer females with feminine faces (Cunningham, 1986; Rhodes, 2006).

3.5 CONCLUSION

Craniofacial bone structure determines subjective impressions about the sex of a male face. This is not the case in females. In females, the mediation of the relationship between facial features and sex judgments by skull revealed only when adjusting facial features for body fat. Body fat has a slight positive effect on correct recognition of male faces but a medium to large negative effect on correct recognition of female faces. This substantial role of fat in females might explain the lack of mediation of the relationship between facial features and sex judgments by skull when body fat is not taken into account.

Next chapter will explore whether development of skull features might be influenced by exposure to prenatal androgens.

CHAPTER 4

Identifying craniofacial features associated with prenatal exposure to androgens and testing their relationship with brain development

4.1 INTRODUCTION

According to the organizational hypothesis, prenatal period is a critical window when androgens impact the development of both reproductive and non-reproductive tissues (Phoenix et al., 1959). In human studies, putative effects of prenatal androgens on brain and behavior have been demonstrated using a variety of approaches, including direct (testosterone level in amniotic fluid [Van de Beek et al., 2004] or umbilical-cord blood [Sakai et al., 1992]) and indirect (the ratio of the length of the 2^{nd} and 4^{th} fingers [2D:4D ratio]; reviewed in Hönekopp et al., 2007) measures. Given the scarce opportunities for measuring directly androgen levels in amniotic fluid or umbilical-cord blood, the discordant-sex twin design has served as an alternative avenue for testing putative effects of prenatal androgens. While female fetus produces androgens only by fetal adrenal glands (Rainey et al., 2004) and as a by-product of corticosteroid production (Tapp et al., 2011), male fetus develops testes in the 7-8 week of gestation (Tapp et al., 2011) and starts producing increasing levels of testosterone

(McIntyre, 2006). Stable testosterone levels have been observed in the amniotic fluid of male fetuses from the earliest measurements taken (15 week of pregnancy; Sarkar et al., 2007). Thus, intrauterine presence of a male (vs. female) co-twin exposes the other twin to higher levels of prenatal androgens and, as such, the discordant-sex twin design allows one to test so-called “testosterone transfer” hypothesis (Peper et al. 2009; Cohen-Bendahan, 2004). Here, we used the discordant-sex twin design to identify a peripheral “signature” of the prenatal exposure to androgens. Given the growing availability of magnetic resonance (MR) images in population-based studies of brain development (Paus, 2013) and the current work on MR-based craniofacial morphometry (e.g. Chakravarty et al. 2011, Liu et al. 2012), we have decided to explore this possibility in the context of craniofacial features.

The concurrent phase of the brain and craniofacial development takes place between the 5th and 13th week of gestation (Diewert et al., 1993; Diewert & Lozanof, 1993). Examples of environmental and genetic perturbations affecting both phenotypes include, respectively, fetal alcohol syndrome (Larkby & Day, 1997) and Down syndrome (Guihard-Costa et al., 2006). Sex differences in the craniofacial morphology have been observed as early as in 6-month old human infants (Bulygina et al 2006). Studies in adults suggested a relationship between 2D:4D ratio and robusticity (Fink et al., 2005; Meindel et al., 2012), dominance, and perceived masculinity (Neave et al., 2003) of the face. Prenatal androgens also appear to influence teeth size: androgenized female monkeys (vs. female controls) had longer and

sharper teeth (Zingeser, & Phoenix, 1978), female-to-male transsexuals (vs. female controls) had larger bucolingual and, to a lesser extent, mesiodistal diameters of the top of the crown (Antoszewski et al., 2009), and females with a twin brother (vs. with a twin sister) had larger mesiodistal and, to a lesser extent, bucolingual diameters of the top of the crown (Dempsey et al., 1999).

Here we use the discordant-sex twin design to identify possible effects of prenatal androgens on craniofacial morphology in pre-pubertal children (Study 1). Subsequently, we test the validity of such same-sex vs. opposite-sex differences, or the craniofacial "signature" of prenatal androgens, by examining its relationship with brain size in a large sample of adolescent females born as singletons (Study 2).

4.2 STUDY 1: METHODS

4.2.1 Participants

A sample of 119 dizygotic 8-year old twins from Quebec Newborn Twin Study included 63 females (28 with a twin-sister [Same Sex Female; SSF], 35 with a twin-brother [Opposite Sex Female; OSF]) and 56 males (20 with a twin-brother [Same Sex Male; SSM], 36 with a twin-sister [Opposite Sex Male; OSM]). For the same-sex groups (i.e., SSF and SSM), we included in our analyses only one member of each twin pair (chosen at random when both MR images were usable).

4.2.2 Image acquisition

MRI T1-weighted data of the whole head were obtained on a 1.5 Tesla system (Magnetom Vision, Siemens Electric, Erlangen, Germany) using TE = 10ms, TR = 22ms, flip angle = 30 degrees, 160 contiguous slices, matrix size = 224 x 256, 1mm x 1mm x 1mm voxels.

4.2.3 Landmarks and landmark-based variables

All 119 T1-weighted images were registered, using rigid transformation (3 translations, 3 rotations), to a T1-weighted image of one individual chosen at random. This ensured similar orientation of all images for the subsequent placement of craniofacial landmarks. Next, we placed 19 skull landmarks (Figure 1) on each of the 119 images using visualization software Register (http://www.bic.mni.mcgill.ca/ServicesSoftware/HomePage). While T1-weighted MR images provide very good contrast between soft tissues, bone tissue can only be estimated as the black non-tissue space. We placed skull landmarks in those non-tissue spaces that enabled high precision in positioning (e.g. particular teeth, corners of the eye sockets, tip of the chin). This approach allowed us to capture craniofacial features that are independent of the amount of fat (or muscle) in the face. Next, we performed landmark-based registration (3 translations, 3 rotations, 3 scaling, 3 sheering) to remove possible differences in the overall craniofacial size. Finally, we extracted X, Y and Z coordinates of the 19 landmarks for each of the 119 individuals.

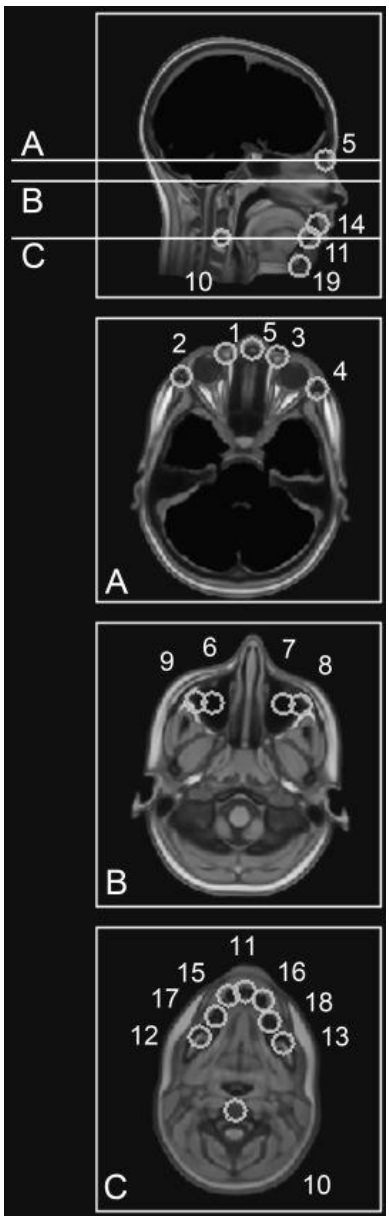


Figure 1: Skull landmarks (19) and their position:

1- Inside corner of the left eye socket;
2- Outside corner of the left eye socket;
3- Inside corner of the right eye socket;
4- Outside corner of the right socket;
5- Bridge of the nose;
6- Middle of the left mandibular sinus;
7- Middle of the right mandibular sinus;
8- Outside of the right mandibular sinus;
9- Outside of the left mandibular sinus;
10- Point around 2nd vertebrae where lower teeth most visible;
11- Front of the lower teeth;
12- Left jaw (unerupted lower 3rd molar)
13- Right jaw (unerupted lower 3rd molar);
14- Front of the upper teeth
15- Left canine;
16- Right canine;
17- Left first molar;
18- Right first molar;
19- Tip of the chin bone.

Landmark coordinates were normalized to a range 0-1 and mean-centered before submitting them to Principal Component Analysis (PCA). In addition, 17 craniofacial distances (Table 1) were calculated as Euclidean distances between the normalized and mean-centered landmark coordinates. Statistical software JMP was used to test the effect of twin group on the principal components (PCs) and landmark distances.

Table 1: Skull landmark distances (17)

Landmark numbers	Landmark distance	Description
1-2	left eye socket length	length of the left eye socket
3-4	right eye socket length	length of the right eye socket
1-3	inside eye corners distance	distance between the inside corners of eye sockets
5-19	nose bridge to chin	distance between nose bridge and tip of chin
5-11	nose bridge to lower front teeth	distance between nose bridge and front teeth of lower jaw
5-14	nose bridge to upper front teeth	distance between nose bridge and front teeth of upper jaw
19-11	chin to lower front-teeth	distance between tip of chin and front teeth of lower jaw (chin height)
6-9	left maxillary sinus	size of the left maxillary sinus
7-8	right maxillary sinus	size of the right maxillary sinus
6-7	maxillary sinuses distance	distance between maxillary sinuses
10-11	spine to lower front-teeth	distance between second vertebrae and front teeth of lower jaw
10-19	spine to chin	distance between second vertebrae and tip of chin
12-13	left-third-molar to right-third-molar	distance between left and right unerupted lower third molars (wisdom teeth)
19-12	left-third-molar to chin	distance between tip of chin and unerupted lower left third molar (wisdom tooth)
19-13	right-third-molar to chin	distance between tip of chin and unerupted lower right third molar (wisdom tooth)
11-12	left-third-molar to lower front-teeth	distance between front teeth of lower jaw and unerupted lower left third molar (wisdom tooth)
11-13	right-third-molar to lower front-teeth	distance between front teeth of lower jaw and unerupted lower right third molar (wisdom tooth)

4.3 STUDY 1: RESULTS

4.3.1 Age

All participants were 8 years old (M=101.1 months, SD=1.03; range: 99.5 – 106.6 months) and their age did not vary as a function of twin group ($F(3,115)=1.13$, $p=0.34$).

4.3.2 Craniofacial features

The PCA identified 10 main PCs describing a total of 78% of variance in the craniofacial features. Four-way ANOVA was carried out to test for differences in PCs loadings across the four groups of twins (SSF, OSF, OSM, SSM). As reported in Table 2, only PC3 skull features showed a main effect of twin-group ($F(3,115)=7.3$, $p=0.0002$ uncorrected; $p=0.002$ corrected for 10 comparisons). Post-hoc analyses showed that SSF group had higher loadings of PC3 skull features than any other group (SSF vs. OSF: $t(61)=3.01$, $p<0.0001$, Cohen's $d=0.76$; SSF vs. OSM: $t(62)=4.35$, $p<0.0001$, Cohen's $d=1.08$; SSF vs. SSM: $t(46)=-2.56$, $p=0.01$, Cohen's $d=0.77$; Figure 2). There were no differences in PC3-loadings across the OSF, OSM, and SSM groups; in particular, OSF did not differ from either OSM ($t(69)=-1.51$, $p=0.14$) or SSM ($t(53)=-0.003$, $p=0.99$). Craniofacial features characteristic for positive values of PC3 are illustrated in Figure 3.

Table 2A: Principle components (PCs) explaining variance in craniofacial features and their means and standard deviations in the four twin groups (SSF, OSD, OSM, SSF*).

PCs and variance explained		SSF (n=28)		OSF (n=35)		OSM (n=36)**		SSM (n=20)	
#	%	M	SD	M	SD	M	SD	M	SD
1	20.7	0.84	3.04	0.29	3.34	-0.57	2.76	-0.55	4.48
2	14.4	-0.62	2.87	-0.64	2.61	0.78	2.57	0.29	2.48
3	9.0	1.50	2.46	-0.17	1.93	-0.84	1.84	-0.17	1.82
4	7.2	-0.11	1.87	-0.12	2.03	-0.04	1.87	0.08	1.82
5	6.0	0.15	1.15	0.07	1.27	-0.38	1.14	-0.47	1.46
6	5.8	-0.13	1.80	0.12	1.60	0.03	1.20	0.44	1.83
7	4.7	0.41	1.79	-0.04	1.47	-0.19	1.36	-0.17	1.22
8	3.7	-0.24	1.51	-0.01	1.57	0.06	1.24	-0.43	1.23
9	3.6	-0.34	1.14	-0.26	1.50	0.26	1.24	0.23	1.35
10	3.1	0.04	1.55	0.20	1.21	0.31	1.87	-0.81	1.27

The first 10 PCs could explain 78% of variance, the first 4 PCs could explain 51% of variance; PC1 – longer nose and prominent chin; PC2 – shorter nose, greater distances between inside corners of the eye sockets; PC3 – prominent chin, wide lower jaw, smaller eye sockets distance; PC4 – smaller distance between mandibular sinuses, greater distance between inside corners of the eye sockets.

*SSF = females with a same sex co-twin, OSF = females with opposite sex co-twin, OSM = males with opposite sex co-twin, SSM = males with same sex co-twin

** PC1-PC3 based on all 119 twins, PC4-PC10 based on 118 twins (1 outlier from OSM group excluded)

Table 2B: Results of four-way ANOVA exploring the effects of twin group (SSF, OSF, OSM, SSM) on craniofacial features (PCs).

PC	F(3,115)	Uncorrected p	Corrected p *
1	1.22	0.31	ns
2	2.32	0.08	ns
3	7.30	0.0002	0.002
4	0.06	0.98	ns
5	1.76	0.16	ns
6	0.53	0.66	ns
7	1.01	0.39	ns
8	0.67	0.57	ns
9	1.68	0.18	ns
10	3.95	0.01	ns

* ns=not significant after correcting for 10 multiple comparisons

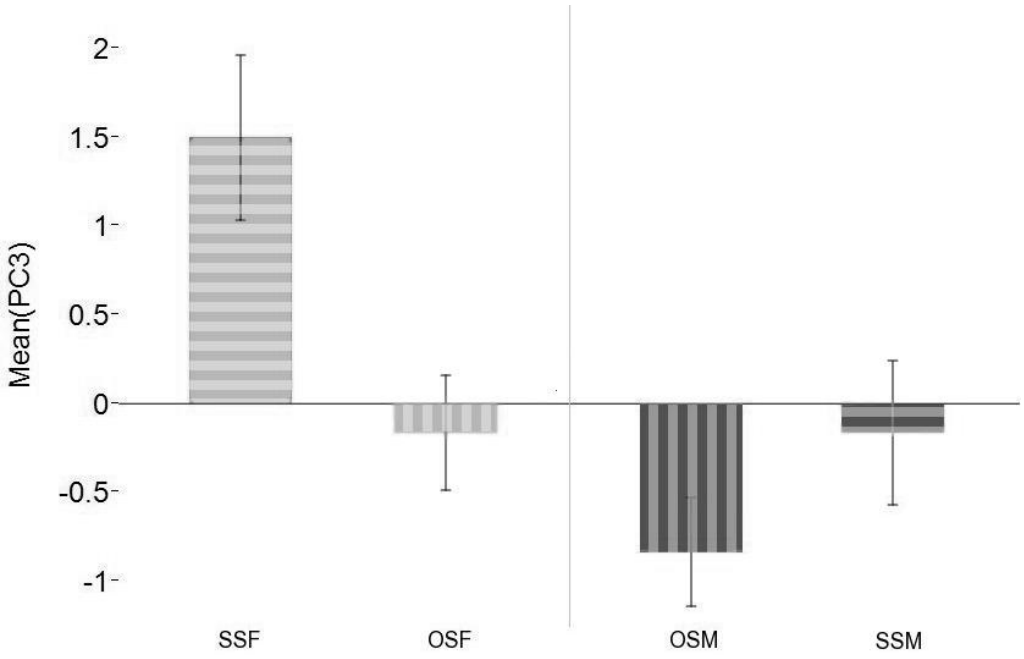


Figure 2: Effect of twin group on presence of PC3 features (SSF = females with female co-twin, OSF = females with male co-twin, OSM = males with female co-twin, SSM = males with male co-twin). SSF had more PC3 features than any other twin group and the effect size of all three comparisons was large.

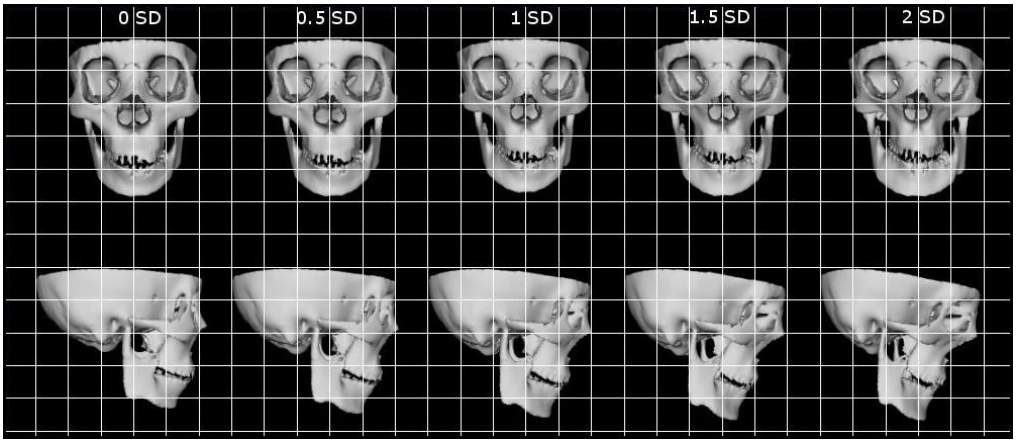


Figure 3: Simulation of the PC3 craniofacial features. The first column of images represents craniofacial features characteristic for the mean values of PC3 further columns of images represent the positive range of PC3 values (0.5 up to 2 SD from the mean) that reflect craniofacial features characteristic for low exposure to prenatal testosterone. Please notice the increasing width and length of the jaw in the upper and lower row of the images, respectively.

Next, we determined which of the 17 craniofacial distances were related to PC3 features and which of them showed a SSF vs. OSF difference. Correlations between the 17 craniofacial distances and PC3 features identified the same eight PC3-related distances in both the whole sample of twins (Table 3) and the female twins only. We then evaluated differences between OSF and SSF and found that five of these eight PCs showed an effect of co-twin's sex (Table 4). Overall, SSF (vs. OSF) had smaller distance between the inside corners of eye sockets, larger distance between left and right third molars of the lower jaw, larger "left-third-molar to chin", "left-third-molar to lower front-teeth", and "right-third-molar to lower front-teeth" distance. This is consistent with the direction of a simple sex difference: females (vs. males) had also smaller distance between the inside corners of eye sockets ($t(116)=3.37$, $p=0.01$, Cohen's $d=-0.63$), larger distance between left and right third molars of the lower jaw ($t(116)=-2.00$, $p=0.05$, Cohen's $d=0.37$), and larger "left-third-molar to chin" ($t(116)=-2.75$, $p=0.007$, Cohen's $d=0.5$), "left third- molar to lower front-teeth" ($t(116)=-4.65$, $p<0.0001$, Cohen's $d=0.86$), and "right-third-molar to lower front-teeth" ($t(116)=-3.88$, $p=0.0002$, Cohen's $d=0.72$) distance.

Table 3: Correlations between the 17 skull landmark distances and PC3 features among all twins.

PC	Distance	correlation	p-value	corrected p-value
PC3	left-third-molar to lower front-teeth	0.62	0.00	0.00
PC3	left-third-molar to right-third-molar	0.56	0.00	0.00
PC3	left-third-molar to chin	0.56	0.00	0.00
PC3	maxillary sinuses distance	-0.49	0.00	0.00
PC3	right-third-molar to lower front-teeth	0.46	0.00	0.00
PC3	Inside eye corners distance	-0.44	0.00	0.00
PC3	right-third-molar to chin	0.43	0.00	0.00
PC3	spine_lower front-teeth	-0.42	0.00	0.00
PC3	nose bridge to lower front-teeth	-0.13	0.16	2.75
PC3	left maxillary sinus	0.12	0.19	3.27
PC3	right eye length	0.11	0.24	4.08
PC3	nose bridge to upper front-teeth	-0.09	0.34	5.74
PC3	nose bridge to chin	0.07	0.46	7.86
PC3	chin to lower front-teeth	0.06	0.50	8.42
PC3	left eye length	-0.03	0.72	12.31
PC3	spine to chin	-0.02	0.86	14.58
PC3	right maxillary sinus	0.01	0.89	15.20

Table 4: Differences in PC3-related distances between females with female (SSF) and male (OSF) co-twin

Distance	OSF M	OSF SD	SSF M	SSF SD	t-value	p-value	Cohen's d
left-third-molar to lower front-teeth	50.04	3.58	52.34	4.01	2.37	0.02	-0.61
left-third-molar to right-third-molar	53.78	4.46	56.47	3.27	2.62	0.01	-0.69
maxilary sinuses distance	41.85	2.89	41.1	2.69	-1.04	0.3	0.27
right-third-molar to lower front-teeth	50.79	3.22	52.57	3.40	2.08	0.04	-0.54
left-third-molar to chin	48.63	4	51.18	4.1	2.45	0.02	-0.63
spine to lower front-teeth	83.01	2.97	82.08	2.55	-1.29	0.2	0.34
right-third-molar to chin	50.32	3.88	52.34	4.45	1.9	0.06	-0.48
inside eye corners distance	30.9	2.45	28.96	2.43	-3.08	0.003	0.8

4.4 STUDY 2: METHODS

In this experiment, we applied the model of PC3 features in an independent dataset of female adolescents and tested for the presence of a relationship between the PC3-related craniofacial features and prenatal androgens using brain size, a known correlate of prenatal androgens (Peper et al., 2009). We hypothesized that the presence of

PC3-related features, an indicator of an *absence* of prenatal androgens (for PC3: SSF > OSF) would predict smaller brain size.

4.4.1 Participants

A total of 462 female adolescents born as singletons (age range = 12 to 18 years; M=180.02 months, SD=22.61) were recruited in the context of Saguenay Youth Study (SYS), a community-based study of typically developing adolescents (Pausova et al, 2007).

4.4.2 MRI data acquisition

T1-weighted head MR images were acquired on a Philips 1T scanner (TR=25ms, TE=5ms, flip angle = 30°, 140-160 slices, resolution 1x1x1mm).

4.4.3 Brain size

Structure-wise volumes were estimated using a pipeline based on a modified version of the ANIMAL algorithm (Collins et al., 1995). The pipeline starts with the correction of intensity inhomogeneities due to radio frequency (RF) field uniformity (Sled et al., 1998) and slice-wise intensity normalization using the median of slice-wise intensity ratios (Zijdenbos et al., 2002). This is followed by linear and nonlinear registration (using the ANTs algorithm (Avants, et al., 2008)) to a template based on the nonlinear average of 808 MRIs (Grabner et al., 2006) acquired from the SYS adolescents. Finally, we used nonlinear

registration to project a brain mask (including cerebellum and brain stem) to native space of SYS adolescents to estimate brain size.

4.4.4 Craniofacial features and projection of facial signature

The process of deriving craniofacial features was accomplished as follows. First, we created a population-based average of all SYS head MR images and normalized each head MRI to this average using 12-parameter registration (3 translations, rotations, scales, and sheers). Next, we placed 19 landmarks (Figure 1; the same landmarks as used in Study 1) on this average image at anatomically defined locations of the skull. Then, we calculated non-linear transformation matrix necessary to register each individual's T1-weighted image to the population-based average. Finally, the craniofacial landmarks were warped back to each participant's head MRI using the inverse of the nonlinear transformation. This step provided a set of landmarks (and relevant euclidean distances; Table 1) for each participant's skull.

Next, we transferred the model of twin PC3 to the SYS average head MRI as follows. First, we calculated the mean X, Y and Z coordinates for each of the 19 landmarks in the twin dataset (these coordinates were the output of the 12-parameter landmark-based registration and were therefore adjusted for head size) and registered them linearly (3 translations, 3 rotations) to the 19 landmarks on the SYS population-based average head MRI. Next, we multiplied these 19 skull coordinates by the PC3 weights and thus created a set of 19 PC3-

like coordinates. Finally, we calculated the difference between the individual skull features (19 skull coordinates for each SYS participant) and this model of PC3-like coordinates using the root mean square error (RMSE) formula. Individuals with smaller RMSE values, had more skull features of PC3, and had thus likely less exposure to prenatal androgens (PC3: SSF>OSF).

4.5 STUDY 2: RESULTS

Relationship between PC3 features (RMSE) and brain size was explored in a sample of 452 females (10/462 participants were excluded due to failing the image-processing pipeline for brain size). As predicted, linear regression showed a positive relationship between RMSE and brain size ($t(451)=2.89$, $p=0.004$; $R^2=0.02$).

In order to identify craniofacial distances with the strongest relationship with brain size, we examined six PC3-related distances that showed a difference between the SSF and OSF group. Three of these distances correlated with brain size and survived correction for 6 multiple comparisons: “right-third-molar to chin” ($r=-0.29$, $p<00001$), “left-third-molar to chin” ($r=-0.24$, $p<0.0001$), and “left-third-molar to lower front-teeth” ($r=-0.14$, $p=0.002$).

Finally, we calculated an average of the two symmetrical distances that showed relationship with brain size (“right-third-molar to chin” and “left-third-molar to chin”). Linear regression showed a negative relationship between the average of these distances and brain size

(t(451)=-6.41, p<0.0001). This fine-tuned craniofacial signature of prenatal androgens explained 8% variance in brain size.

4.6 DISCUSSION

We used twin design and head MR images to study the possible effect of prenatal androgens on craniofacial features. Females with a female co-twin differed from all the other twin groups that were exposed to prenatal androgens (OSF, OSM, SSM) and the effect size was large (Cohen's d ~ 0.8) in all three contrasts (SSF vs. OSF, SSF vs. OSM, SSF vs. SSM). Same-sex females had higher loadings of PC3 skull features, and thus shorter distance between the inside corners of the eye sockets, larger distance between left and right third molars of the lower jaw, and larger distance between the left third molar and lower front-teeth, right third molar and lower front-teeth, left third molar and tip of the chin, and right third molar and tip of the chin.

In order to verify the relationship between these craniofacial features and prenatal androgens, we used a large independent dataset of female adolescents (singletons) to explore the relationship between skull features and brain size, a known correlate of prenatal androgens (Peper et al., 2009). We confirmed our prediction, namely that PC3-related features would be negatively related to brain size: the set of PC3-related features could explain 2% variance in brain size, and the mean distance between the left third molar to the tip of chin and the right third molar to the tip of chin could explain 8% variance in brain size.

Sex differences in face shape appear already in the first year of life (Bulyina et al., 2006). It is possible that this sexual dimorphism in craniofacial development might be related to prenatal androgens. Male infants have a relatively larger and more globular frontal bone, smaller face, and a more flexed cranial base than female infants (Bulygina et al., 2006). While a comparison of these findings with our results from Study 1 is difficult due to the different sets of skull landmarks, we speculate that the larger frontal bone in male vs. female infants (Bulygina et al., 2006) might be consistent with the larger brain size (Peper et al., 2009) and greater distance between the inner corners of eye sockets, characteristic for females with a male vs. female co-twin (Study 1). Craniofacial features that showed an influence by prenatal androgens in our study also correspond to the embryonic development. Testes develop at 7-8 weeks of gestation (Tapp et al., 2011), which is a period characterized by mandibular and maxillary ossification, formation of deciduous tooth buds, and migration of eyes medially (Sperber et al., 2000).

In the twin study (Study 1), we found differences between the same-sex females and each of the other three twin groups, but no difference between the opposite-sex and same-sex males. This is consistent with the effect of co-twin's sex on teeth size reported by Dempsey et al (1999). It seems that the effect of prenatal androgens on the skull appears at certain level of prenatal androgens but does not follow a simple (linear) dose response. Females produce very little endogenous testosterone and therefore gestation with a male co-twin

has a relatively greater effect on females compared with males (Tapp et al., 2011).

The large effect of twin group on PC3 (Cohen's d = 0.76 in SSF vs. OSF, 0.77 in SSF vs. SSM, and 1.08 in SSF vs. OSM) is consistent with van Anders et al.'s (2006) study about the effect of co-twin's sex on digit ratio. The small (2%) effect of PC3-related features (RMSE) on brain size in females from Study 2 is consistent with Peper et al. (2009) who reported a small difference in brain size between same- vs. opposite-sex females (Cohen's d = 0.36, which is R^2=0.03). The fine-tuned facial signature (i.e. mean of two distances: the "left-third-molar to tip of chin" and "right- third-molar to tip of chin") explained 8% in brain size, suggesting that the facial signature – estimated in singletons - might have a comparable (or even greater) power as the twin design to study effects of prenatal androgens.

Overall, these findings suggest that prenatal androgens did leave their signature in the face and that this facial signature might be used, similarly to digit ratio, as an indirect index of exposure to prenatal androgens. Given the widespread availability of T1-weighted head MRIs, an MR-based facial signature might be easily accessible to many researchers interested in the effects of prenatal androgens.

Moreover, it is possible that facial signature might be a more accurate indicator of prenatal androgens than digit ratio. While sex differences in the levels of prenatal testosterone (measured directly) are large (Cohen's d=1.9; Knickmeyer et al., 2005; Van de Beek et al.,

2004), sex differences in digit ratio are rather small (Cohen's d = 0.2 in Manning et al., 2007 and 0.3 in Manning et al., 2004). Given the medium effect size of the sex differences in PC3 (Cohen's d=0.56), the facial signature may provide a better - albeit still indirect - index of prenatal exposure to androgens. Moreover, predicting exposure to prenatal androgens from facial signature might be more straightforward than testing both left and right digit ratio and then choosing which one is able to show a relationship with prenatal androgens. The inconsistency of right vs. left digit ratio as an index of prenatal androgens was reviewed, for example, in McInthyre (2006). Reliability of digit ratio as an indirect index of prenatal androgens has been also questioned because several studies failed to support the presence of relationship between digit ratio and direct measures of prenatal androgens. Findings about the relationship between digit ratio and prenatal androgens (using dizygotic twin design [Van Anders & Verhorn 2006]; females with congenital andrenal hyperplasia vs. controls [Brown et al., 2002; Okten et al., 2002]; females with complete androgen insensitivity syndrome vs. controls [Berenbaum et al., 2009]), were not supported by Lutchmaya et al. (2004) who had access to direct measures of prenatal sex hormones from amniotic fluid but found no relationship between prenatal androgens and digit ratio, only a relationship between digit ratio and the ratio of foetal testosterone and foetal estrogen. Buck et al. (2003) who studied digit ratio in females with congenital adrenal hyperplasia and control. and had more than double sample size than either Brown et al (2002) or Oktern et al. (2002), reported no differences in digit ratio.

An animal study found a relationship between digit ratio and maternal corticosterone but not maternal testosterone, suggesting that digit ratio might reflect maternal stress rather than levels of prenatal androgens. Future research is needed to clarify the relationship between facial signature and prenatal androgens, for example by examining the MR-based facial signature in datasets with available amniotic fluid data (Lombardo, et al., 2012).

4.7 CONCLUSION

We used a cohort of 8-year old dizygotic twins to study the relationship between prenatal androgens and craniofacial shape. Head MR images enabled us to describe variability in skull features among these twins. Females with a female co-twin showed skull features that distinguished them from all other twin groups exposed to at least some levels of prenatal androgens. In order to verify the existence of the relationship between prenatal androgens and skull features, we studied relationship of this facial signature with brain size, a known correlate of prenatal androgens, in a large independent sample of female adolescents born as singletons. Facial signature predicted 2% and the mean distance between chin and sides of the jaw predicted 8% variability in brain size. We conclude that this signature of prenatal androgens in the face might be used in future studies as an alternative to digit ratio to study the role of prenatal androgens on brain and disease risk.

Next chapter will focus on postnatal sex hormones and study their influence on brain response to faces and eye-movements scanning the face.

CHAPTER 5

Hormonal contraceptives, menstrual cycle and brain response to faces

5.1 INTRODUCTION

Robust sex differences exist in face perception. A meta-analysis exploring sex differences in the development of facial-emotion recognition showed that the origins of this female advantage might be present already in the first year of life (McClure, 2000). Both behavioral and neuroimaging studies agree that the recognition of facial emotions and the related neural architecture continue to develop throughout childhood and adolescence (Herba & Philips, 2004), and that the peak in emotion recognition is reached in young adulthood (Sullivan et al., 2007). While the presence of sex differences in emotion recognition in the first year of life suggests a possible role of genes and/or prenatal exposure to androgens, it seems that postnatal sex hormones might play a role in face perception as well (e.g. Derntl et al., 2008). Sex differences in face perception and emotion recognition vary in their magnitude. Women outperform men in face detection, the very first stage of face perception (Cohen's d=0.91; McBain et al., 2009). Women also outperform men in accuracy and speed of recognizing emotions in faces (Cohen's d=0.3-0.4; Hall & Matsumoto, 2004; Hampson et al.,

2006; Kirouac & Dore, 1985; Rahman et al., 2004), including situations when emotions are displayed at lower intensities (Montagne et al., 2005). This female advantage in emotion recognition was predicted by both dwell time and number of fixations to the eyes when scanning the face (Hall et al. 2010).

Sex differences in neuroimaging findings parallel these behavioral differences in face perception. Schulte-Rüther et al. (2008) reported that during perception of emotional faces, women (vs. men) had higher blood oxygenation level-dependent (BOLD) response in face processing regions, such as the right superior temporal sulcus (STS). A large study of typically developing adolescents (518 males, 592 females) found that females watching emotionally ambiguous faces had a stronger BOLD response than males in a number of cortical regions, including the fusiform face area (Tahmasebi et al., 2012), a region selectively involved in the perception of faces (Kanwisher et al., 1997). As mentioned above, sex hormones might, in part, explain the presence of the above sex differences. As estrogen enhances performance on sexually dimorphic tasks that favor women (e.g. verbal memory) and impairs performance on tasks that favor men (Sandres et al., 2002; Hampson, & Kimura, 1988), estrogen might also modulate face perception. Only a handful of studies have investigated this topic. Miyahira et al. (2000) found that sex differences in exploratory eye movements emerge alongside hormonal changes following puberty and disappear following menopause, thus suggesting that some (general)

aspects of visual information processing might be regulated by sex hormones.

In this study we wanted to explore whether variations in sex hormones influence brain response to faces. Menstrual cycle and the use of oral contraception (OC) provide an opportunity to study such effects given the fluctuations in sex hormones associated with both. Since estrogen and progesterone levels are higher during mid-cycle than menstruation, we hypothesized that women would have higher BOLD response to faces in the face processing network in general, and in the FFA in particular, during mid-cycle. We also predicted that women taking OC, and therefore having higher levels of estrogen and progesterone due to exogenous hormones, would have higher BOLD response to faces compared with freely cycling women. In addition, we also explored whether the duration of OC use modulates the FFA response in a dose-related manner.

We replicate these findings in a sample of adolescent females, thus exploring whether the effects of OC on FFA BOLD response can be detected as early as adolescence. Given the expected dose effect of OC use-duration, we predicted that the effect of OC on BOLD response in FFA would be most likely smaller in adolescents compared with adult women. Finally, we tested whether such effects of sex hormones on brain response to faces might be reflected in the pattern of eye movements scanning the face. Therefore, we conducted an eye-tracking study using the same face stimuli in another sample of women;

we predicted that women using OC might show longer fixations to the eye region.

5.2 METHODS AND MATERIALS

5.2.1 Experiment I: functional MRI in young women

5.2.1.1 Participants

Twenty healthy women between 18 and 29 years of age were recruited at the University of Nottingham (UK): 10 freely cycling and 10 taking OC. The three types of OC used (Ovranette, Microgynon, Femodene) had the same amount of Ethinyl Estradiol (30 µg) and slightly different amounts of progestin (150 µg of Levonogestrel in Ovranette and Microgynon and 75 µg of Gestogene in Femodene). Duration of OC varied between 3 and 53 months (*M*=20.56 months, *SD*=16.89). There was no significant difference between the age of freely cycling women (*M*=20.44 years, *SD*=2.69) and women taking OC (*M*=22.0 years, SD=3.26; $t_{(18)}$=1.12; *p*=0.28). Participants' consent was obtained according to the Declaration of Helsinki and approved by the Ethics Committee of the Medical School at the University of Nottingham.

5.2.1.2 Design

In order to schedule the MR visits, each participant filled a brief questionnaire about their menstrual cycle (average length of cycle, average length of menstruation, date of last menstruation, brand and

duration of OC use), and tracked basal body temperature for one menstrual cycle prior to scanning. This information was taken into account when scheduling the subsequent visits. Basal body temperature was measured orally right after awakening so that the ~approximate 0.5-C increase in temperature just prior mid-cycle could be detected. All 20 participants provided a blood sample and took part in a structural MRI session four times: (1) menstruation (day 5±2 days); (2) follicular phase (day 11±2 days); (3) mid-cycle (day 15±2 days); and (4) late luteal phase (day 28±2 days). The structural MRI session included T1-weighted images (T1W), magnetic transfer ratio (MTR), and diffusion tensor imaging (DTI). In addition, fMRI data were collected twice: at menstruation and mid-cycle. Phases of the menstrual cycle were counterbalanced across the two visits. At the first visit, six women from each group (pill, no pill) were scanned during menstruation and four women from each group were scanned during mid-cycle (and *vice versa* at the second visit).

5.2.1.3 Sex hormones

Blood samples were collected between 9 am and 10:30 am and serum was analyzed at clinical haematology laboratory at the Queen's Medical Centre, Nottingham, UK. Levels of 17-beta oestradiol were obtained via the ADVIA Centaur Estradiol assay (Siemens) and levels of progesterone via ADVIA Centaur Progesterone assay (Siemens). Both immunoassays used direct chemiluminescent technology and the ADVIA Centaur System (Siemens).

The ADVIA Centaur estradiol assay is a competitive immunoassay method employing a sheep anti-E2 monoclonal antibody labeled with acridium ester. Functional sensitivity is 0.019 ng/ml and specificity is high. Cross reactivity with estradiol derivatives, estriol, estrone, testosterone, and testosterone derivatives was less than 0.5 %. Cross reactivity with ethinyl estradiol was 0.037%. Analytical measurement range was 0.0118-3 ng/ml.

The ADVIA Centaur progesterone assay employs acridinium ester-labeled mouse monoclonal anti-progesterone antibody in the Little Reagent. Sensitivity of the ADVIA Centaur Progesterone assay was 0.21 ng/ml and the specificity was high. Cross-reactivity was 0.95% for corticosterone, 0.46% for pregnenolone, 0.31% for 17-alpha-hydroxyprogesterone, 0.08% for 11-deoxycorticosterone and was not detectable for cortisol, testosterone, aldosterone, androstenediol, 11-deoxycortisol, danazol, prednisolone, 17-beta-estradiol, estrone, estriol, clomiphene, and bromocryptine. Analytical measuring range was 0.21 ng/ml to 60 ng/ml.

5.2.1.4 Magnetic resonance imaging: acquisition

Scanning was carried out on a 1.5T Philips scanner. We acquired T1-weighted 1-mm isotropic images and blood oxygenated level-dependent (BOLD) T2*-weighted gradient-echo, echo-planar images (EPI; matrix size 64x64, TE=50 ms, TR=3,000 ms, 180 volumes, voxel size 4x4x4 mm^3). Each EPI image covered the whole brain and consisted of 32 axial slices.

5.2.1.5 Functional paradigm: the face task

Participants passively viewed black–and-white videoclips of faces with ambiguous (e.g. twitching their nose, blinking their eyes, opening their mouth) and angry expressions (Grosbras, & Paus, 2006). In both videoclips the gaze of the actor is direct and forward. This paradigm is identical to the one used in the Tahmasebi et al. (2012) study of typically developing adolescents. Each condition/block lasted for 18 s and included 7-8 videoclips of faces with either ambiguous or angry expression. Non-biological motion (moving circles) was used as a control condition. Henceforth, the term “angry condition” refers to the “angry-control” contrast and the term “ambiguous condition” refers to the “ambiguous-control” contrast.

5.2.1.6 Functional MR images: preprocessing and analysis

All brain imaging data were processed with FMRI Expert Analysis Tool (FEAT), FSL version 4.1. (FMRIB Software Library, www.fmrib.ox.ac.uk/fsl). Pre-processing consisted of motion correction using MCFLIRT, spatial smoothing using Gaussian kernel of FWHM 8mm, and high-pass filter of 100. Functional MR images were registered to each participant’s T1-weighted images that were, in turn, registered to the standard space (152-MNI brain, 2mm) using non-linear registration.

Outcomes of nonlinear registration and MCFLIRT motion-correction were checked. Imaging data affected by translations larger

than 2 mm and rotations larger than 2 degrees (0.035rad) were flagged; this was the case for two women in each group (one session each). For these two images, a confounder variable identifying volumes of no interest was created by fsl_motion_outliers (FSL 4.1.1), added to the design of first level analysis, and new outputs of first level analysis for these two subjects were created.

Functional data were analyzed using a 2 (pill: yes, no) by 2 (phase: menstruation, mid-cycle) analysis of variance (ANOVA). Cluster-based thresholding was used to correct for multiple comparisons: First, images were thresholded voxelwise at $z=1.7$. Second, random field theory was applied to identify clusters that are big enough for the $z=1.7$ to ensure an overall (family-wise) $p<0.05$. The minimal cluster size of the group analysis output (resampled to 2x2x2 mm) for this p-threshold was 252 voxels.

In order to explore the effect of OC duration on mean BOLD response in FFA, the voxelwise approach was complemented by a region-of-interest analysis focusing on the FFA. Similar to Tahmasebi et al. (2012), masks for right and left FFA were created separately for ambiguous and angry condition in the following way: First, we created thresholded images of *z*-statistics ($z>2.3$) for the following four groups: (1) OC during menstruation; (2) OC during mid-cycle; (3) freely cycling during menstruation; and (4) freely cycling during mid-cycle. Second, an intersection with Tahmasebi et al.'s (2012) FFA mask was created for each of these four images. Third, a union of these four intersections

was created. This was repeated separately for the ambiguous and angry contrasts, respectively (Figure 1A). Mean BOLD response in these unionized masks was calculated by Featquery (FSL 4.1.1) and its relationship with OC duration was assessed.

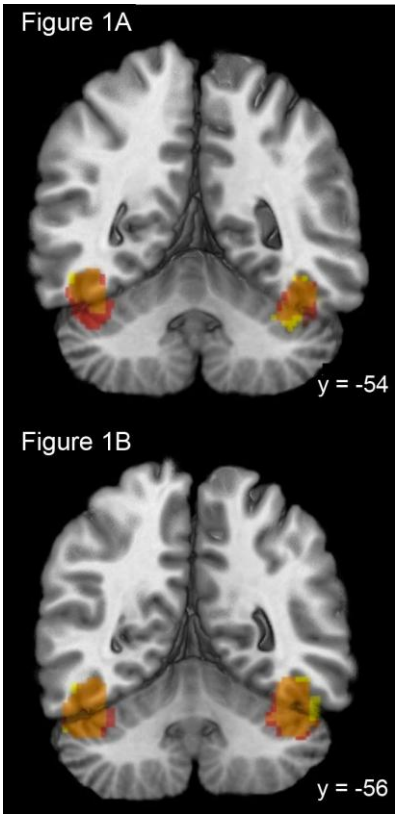


Figure 1A: Experiment I – Unionized FFA masks created for angry (red) and ambiguous (yellow) condition. (Their intersection is in orange.)

Figure 1B: Experiment II – Unionized FFA masks created for angry (red) and ambiguous (yellow) condition. (Their intersection is in organge.)

5.2.2 Experiment II: functional MRI in adolescent girls

5.2.2.1 Participants

A total of 110 adolescent girls were recruited in the context of a European multi-site study *Imagen* (Schumann et al., 2010). All girls were between 13.5 and 15.5 years of age and post-menarcheal (94 in Tanner stage 4, 16 in Tanner stage 5). Local ethics boards approved

the study protocol; the parents and adolescents provided written informed consent and assent, respectively. While the whole *Imagen* study contains data for 2,117 adolescents (1041 females), only 55 female adolescents were taking OC. We matched these 55 adolescent girls taking OC with 55 freely cycling adolescent girls by age (in months), Tanner pubertal stage, and the acquisition site. Information about the type of OC and the length of use was not collected. Possible group differences in personality traits were assessed with NEO-FFI (Costa & McCrae, 1989).

5.2.2.2 Magnetic resonance imaging: acquisition & task

Scanning was carried out on 3T scanners from four different manufacturers (Siemens at 4 sites, Philips at 2 sites, General Electric and Bruker at 1 site, respectively). This report utilizes T1-weighted images (TE=2.8ms, TR=2300ms, voxel size 1x1x1mm^3) and blood oxygenated level-dependent (BOLD) T2*-weighted gradient-echo, echo-planar images (EPI; matrix size 64x64, TE=30ms, TR=2200ms, 160 volumes, voxel size 3.4x3.4x3.4 mm^3). Participants viewed the same stimuli as in Experiment I, namely the black–and-white videoclips of faces with ambiguous and angry expressions (Grosbras, & Paus, 2006).

5.2.2.3 Functional paradigm: preprocessing and analysis

Preprocessing steps were identical to those of Experiment I, with the exception that spatial smoothing used Gaussian kernel of FWHM 7mm (in accordance with the voxel size of the raw images). No imaging

data were affected by head motion larger than 2 mm (translations larger than 2 mm and/or rotations larger than 2 degrees (0.035rad)).

Voxelwise analysis used an unpaired t-test design to explore the effect of pill on BOLD response to faces. Similarly to Experiment I, the correction for multiple comparisons was done using cluster-based thresholding: First, voxelwise images were thresholded at z=1.7, and second, random field theory was applied to identify clusters that are big enough for the z=1.7 to ensure an overall (family-wise) $p<0.05$.

A region-of-interest analysis focused on the FFA. As described above (Experiment I), masks for right and left FFA were created separately for ambiguous and angry condition in the following way: First, we created thresholded images of *z*-statistics ($z>2.3$) for the following two groups: (1) OC and (2) freely cycling girls (n=55 for each group). Second, an intersection with Tahmasebi et al.'s (2012) FFA mask was created for each of these two images. Third, a union of these two intersections was created. This was repeated separately for the ambiguous and angry contrasts, respectively (Figure 1B). Mean BOLD response in these unionized masks was calculated by Featquery (FSL 4.1.1).

5.3 RESULTS

5.3.1 Experiment I: functional MRI in young women

Information about the length of menstrual cycle and length of menses for both pill and no pill group are provided in Table 1A.

Table 1A: Length of menstrual cycle and length of menses in pill and no pill group.

	Freely Cycling	Pill
Length of menstrual cycle (days)	M=26.9 SD=2.46	M=28 SD=0
Length of menses (days)	M=5.35 SD=1.11	M=4.65 SD=1.33

Table 1B: Levels of estrogen and progesterone (means and standard deviations) during menstruation and mid-cycle in women from the pill and no pill group.

	Freely cycling, Menstruation	Freely cycling, Mid-cycle	Pill, Menstruation	Pill, Mid-cycle
Estradiol (pmol/L)	M=144.67, SD=65.23	M=597.68, SD=479.32	M=120.71, SD=96.32	M=88.63, SD=42.5
Progesterone (nmol/L)	M=3.87, SD=1.23	M=12.04, SD=12.43	M=2.95, SD=1.07	M=2.63, SD=1.0

5.3.1.1 Sex hormones

Levels of estrogen and progesterone (means and standard deviations) during menstruation and mid-cycle in women from the pill and non-pill group are provided in Table 1. For estradiol, a two-way repeated measures ANOVA showed main effects of Pill ($F_{(1,18)}$=11.59, p<0.01, Cohen's d=-0.9) and Phase ($F_{(1,18)}$=7.23, p=0.01, Cohen's d=-0.68), and a Pill*Phase interaction ($F_{(1,18)}$=9.6, p<0.01). Student-t post-hoc analyses showed that estradiol levels were higher in freely cycling women during mid-cycle compared with menstruation (t=4.09, p<0.002), higher in freely cycling women than in OC women during mid-cycle (t=-

4.6, p<0.001) but not during menstruation (t=-0.22, p=0.83). For progesterone, a two-way repeated measures ANOVA showed main effect of Pill ($F_{(1,18)}$=6.74, p=0.01, Cohen's d=-0.7) and a Pill*Phase interaction ($F_{(1,18)}$=4.55, p=0.04). Student-t post-hoc analyses showed that progesterone levels were higher in freely cycling women during mid-cycle compared with menstruation (t=2.9, p=0.01), higher in freely cycling than in OC women during mid-cycle (t=-3.34, p<0.01), but not during menstruation (t=-0.33; p=0.75).

5.3.1.2 Functional MRI: voxelwise analysis

A two-way ANOVA showed significant main effects of Pill, Phase, and an interaction between the Pill and Phase (Table 2 and Figure 2A & B). In the pill group (vs. no pill group), we observed a stronger BOLD response in the right FFA to both ambiguous and angry faces. When watching angry faces, a stronger BOLD response in right FFA was also observed in the mid-cycle (vs. menstruation) phase.

Table 2A: Voxelwise approach - Main effect of Contraceptive Pill: pill (+) vs. no pill (-) in ambiguous face condition

#	Label	Hemisphere	Lobe	Voxels	X	Y	Z	z value
1.	FFA	Right	Occipital	6024	40	-54	-20	8.62
2.	Middle Temporal	Left	Temporal	9125	-50	-44	0	7.97
3.	Paracentral Lobule	Right	Frontal	3817	2	-30	66	- 4.51

Table 2B: Voxelwise approach - Main effect of Contraceptive Pill: pill (+) vs. no pill (-) in angry face condition

#	LABEL	HEMISPHERE	LOBE	VOXELS	X	Y	Z	Z VALUE
1.	FFA	Right	Occipital	12091	38	-54	-22	7.51

Table 2C: Voxelwise approach - Main effect of Phase: mid-cycle (+) vs. menstruation (-) in ambiguous face condition

#	**Label**	**Hemisphere**	**Lobe**	**Voxels**	**X**	**Y**	**Z**	**z value**
1.	STS	Right	Temporal	4087	50	-48	6	6.4
2.	IFG	Left	Frontal	6100	-50	20	18	4.97
3.	IFG	Right	Frontal	3201	54	20	24	4.63
4.	Ling. Gyr.	Right	Occipital	9613	2	-86	-6	-5.15

Table 2D: Voxelwise approach - Main effect of Phase: mid-cycle (+) vs. menstruation (-) in angry face condition

#	**Label**	**Hemisphere**	**Lobe**	**Voxels**	**X**	**Y**	**Z**	**z value**
1.	FFA	Right	Occipital	6093	40	-62	-26	6.35
2.	IFG	Left	Frontal	2816	-50	34	8	4.86
3.	Middle Temporal	Left	Temporal	3331	-58	-58	4	4.13

Table 2E: Voxelwise approach – Pill by Phase interaction (no pill mid-cycle and pill menstruation > pill mid-cycle and no pill menstruation) in ambiguous face condition

#	**Label**	**Hemisphere**	**Lobe**	**Voxels**	**X**	**Y**	**Z**	**z value**
1.	Middle Occipital	Left	Occipital	8419	-24	-92	12	5.63
2.	Middle Frontal	Left	Frontal	4013	-36	64	10	3.79

Table 2F: Voxelwise approach – Pill by Phase interaction (no pill mid-cycle and pill menstruation > pill mid-cycle and no pill menstruation) in angry face condition

#	Label	Hemisphere	Lobe	Voxels	X	Y	Z	z value
1.	Middle Occipital	Left	Occipital	11893	-24	-90	14	5.87

5.3.1.3 Functional MRI: ROI analysis of the FFA response

A two-way repeated measures ANOVA showed significant effect of Pill use on the mean BOLD response in the right FFA (Figure 3A, Table 3A and B). There was no effect of Phase and no interaction between the Pill and Phase in the left nor right FFA. Mean BOLD response in the right FFA was higher in the pill group (vs. no pill group) in both ambiguous ($F_{(1,18)}$=5.18, p=0.03, Cohen's d=0.69)) and angry condition ($F_{(1,18)}$=4.51, p=0.04, Cohen's d=0.66). Mean BOLD response in the left FFA did not show any significant differences between the pill and no pill group in the ambiguous ($F_{(1,18)}$=1.58, p=0.22) nor angry condition ($F_{(1,18)}$=0.79, p=0.38). These results are consistent with those of the voxelwise analyses reported above.

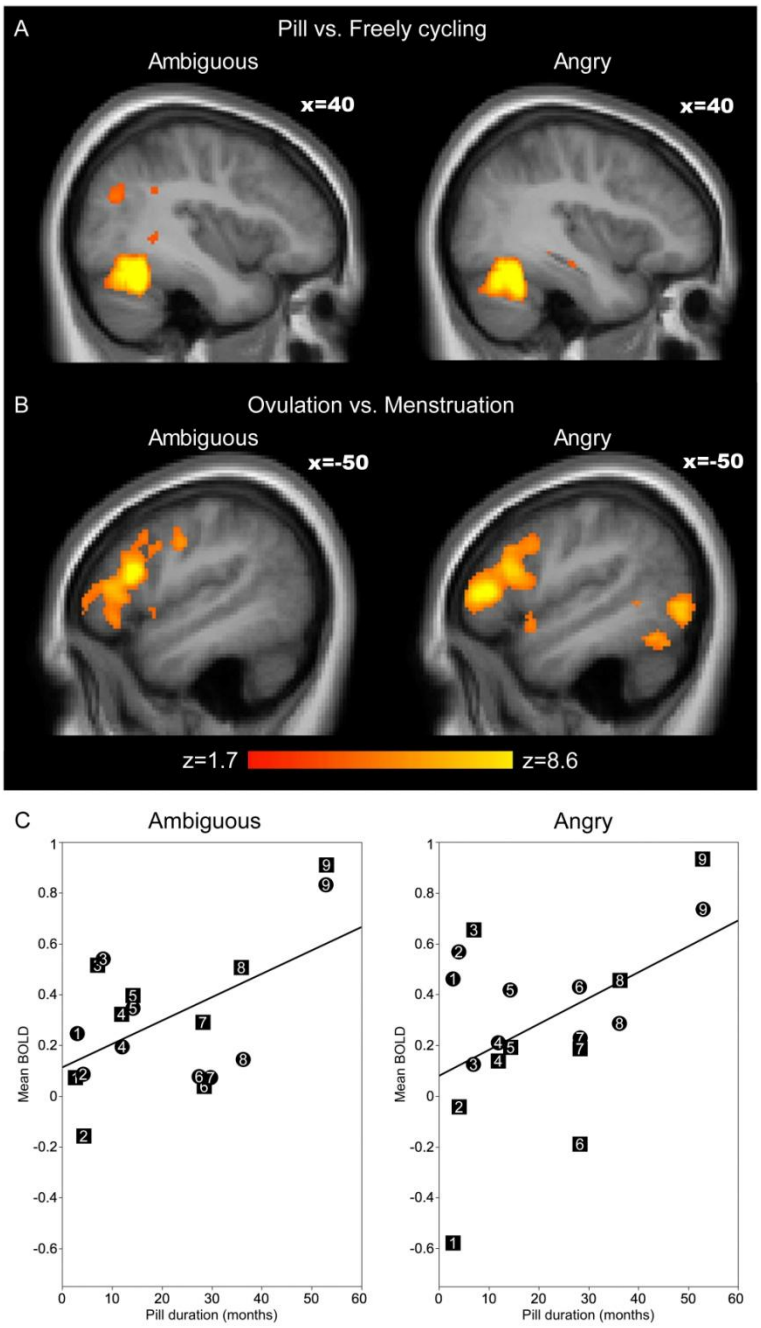


Figure 2A: Effect of oral contraception (pill vs. freely cycling). The left image shows the effect of pill during ambiguous face condition, the right image shows the effect of pill during angry face condition.

Figure 2B: Effect of phase (mid-cycle vs. menstruation). The left image shows the effect of phase during ambiguous face condition, the right image shows the effect of phase during angry face condition.

Figure 2C: Effect of pill duration on mean BOLD response. The left image shows the effect of pill duration on left FFA during ambiguous condition, the right image shows the effect of pill duration on left FFA during angry condition. Each women is represented by a number. Data obtained in menstruation and mid-cycle phase are represented by squares and circles respectively.

Table 3A: Experiment I, ROI approach – Effect of pill and cycle phase on mean BOLD response in FFA (mean and SD).

		Freely, Mid-cycle	Freely, Menstruation	Pill, Mid-cycle	Pill, Menstruation
Right FFA	Ambiguous condition	M=0.22, SD=0.28	M=0.03, SD=0.41	M=0.43, SD=0.22	M=0.25, SD=0.27
Right FFA	Angry condition	M=0.23, SD=0.31	M=0.02, SD=0.52	M=0.47, SD=0.13	M=0.27, SD=0.39
Left FFA	Ambiguous condition	M=0.11, SD=0.32	M=0.23, SD=0.22	M=0.27, SD=0.24	M=0.29, SD=0.32
Left FFA	Angry condition	M=0.15, SD=0.49	M=0.18, SD=0.27	M=0.37, SD=0.19	M=0.15, SD=0.44

Table 3B: Experiment I, ROI approach - Effect of contraceptive pill (pill vs. no pill) on mean BOLD response in FFA, as assessed with 2-way repeated measures ANOVA.

Mask		F	p
Right FFA	Ambiguous condition	5.18	0.03
Right FFA	Angry condition	4.51	0.04
Left FFA	Ambiguous condition	1.58	0.22
Left FFA	Angry condition	0.79	0.38

Table 3C: Experiment I, ROI approach – Results of 2-way repeated measures ANOVA exploring the effect of OC Duration, Phase, and their interaction on FFA BOLD. Here we are reporting the F and R^2 values for the effect of OC Duration on mean BOLD response.

Mask		F	p	R^2	F (corrected for age)	p (corrected for age)	R^2 (corrected for age)
Right FFA	Ambiguous condition	2.48	0.14	0.26	0.94	0.35	0.32
Right FFA	Angry condition	5.7	0.03	0.45	2.66	0.13	0.51
Left FFA	Ambiguous condition	6.2	0.03	0.34	4.87	0.046	0.34
Left FFA	Angry condition	5.1	0.04	0.38	3.72	0.076	0.38

A 3-way repeated measures ANOVA was used to assess the effect of Pill, Cycle, and Condition on BOLD response in the right FFA; it confirmed the significant effect of Pill (F=4.23, p=0.05) and Phase (F=5.06, p=0.04) and showed no effect of Condition (F=0.11, p=0.75) or any condition-related interaction (Pill*Condition: F=0.14, p=0.71; Phase*Condition: F=0.12, p=0.73; Pill*Phase*Condition: F=0.006, p=0.94). A 3-way repeated-measures ANOVA assessing the effects of Pill, Phase, and Condition on BOLD response in the left FFA showed a Phase*Condition interaction (F=5.35, p=0.03) and, as expected, no effect of Pill (F=0.91, p=0.35), Phase (F=0.02, p=0.89), Condition (F=0.12, p=0.73), or any other condition-related interaction (Pill*Condition: F=0.01, p=0.92; Pill*Phase*Condition: F=1.21, p=0.29).

5.3.1.4 Functional MRI: ROI analysis of the amygdala response

Even though the voxelwise results showed no effect of Pill, Phase, or Pill*Phase interaction on BOLD response in amygdala, we have conducted an ROI analysis of the amygdala response upon reviewer's suggestion and we report these results in Appendix 8 now.

5.3.1.5 Sex hormones and FFA BOLD response

We also explored effects of estrogen and progesterone on BOLD response in the right FFA using a three-way repeated-measures ANOVA, which examined the effect of Estrogen (log transformed values), Pill and Phase on BOLD response in right FFA. This analysis revealed a significant three-way interaction in both the ambiguous ($F=23.79$, $p<0.0001$) and angry condition ($F=12.79$, $p=0.001$). We explored this interaction using two-way ANOVA examining the effects of Pill and Estrogen separately for menstruation and mid-cycle. No significant effects were found in mid-cycle. During menstruation, the interaction between Estrogen and Pill was significant (Ambiguous: $F=18.69$, $p=0.0005$; Angry: $F=1.49$, $p=0.01$): the BOLD response increased as a function of estrogen in freely cycling women (Ambiguous: $t(9)=2.8$, $p=0.02$; Angry: $t(9)=2.24$, $p=0.06$) but decreased as a function of estrogen in OC women (Ambiguous: $t(9)=-3.28$, $p=0.005$; Angry: $t(9)=-0.89$, $p=0.4$). No significant effects of progesterone were observed.

5.3.1.6 Duration of OC Use

One woman from the pill group (n=10) had to be excluded from this analysis because of missing information about OC use-duration.

As shown in Table 3C, mean BOLD response in the left FFA increased as a function of OC duration in both the ambiguous ($F_{(1,14)}$=6.2, p=0.03, R^2=0.29) and angry condition ($F_{(1,14)}$=5.1, p=0.04, R^2=0.22; Figure 2C). While mean BOLD response in the right FFA was also increasing as a function of OC duration during the angry condition ($F_{(1,14)}$=5.7, p=0.03, R^2=0.13), no effect was found in the ambiguous condition ($F_{(1,14)}$=2.48 , p=0.14). Note, however, that these results did not survive correction for multiple (four) comparisons.

Given a negative trend between Pill Duration and Age (r=-0.44, p=0.07), we also checked possible effects of Age on the FFA response to faces; none were significant.

5.3.2 Experiment II: functional MRI in adolescent girls

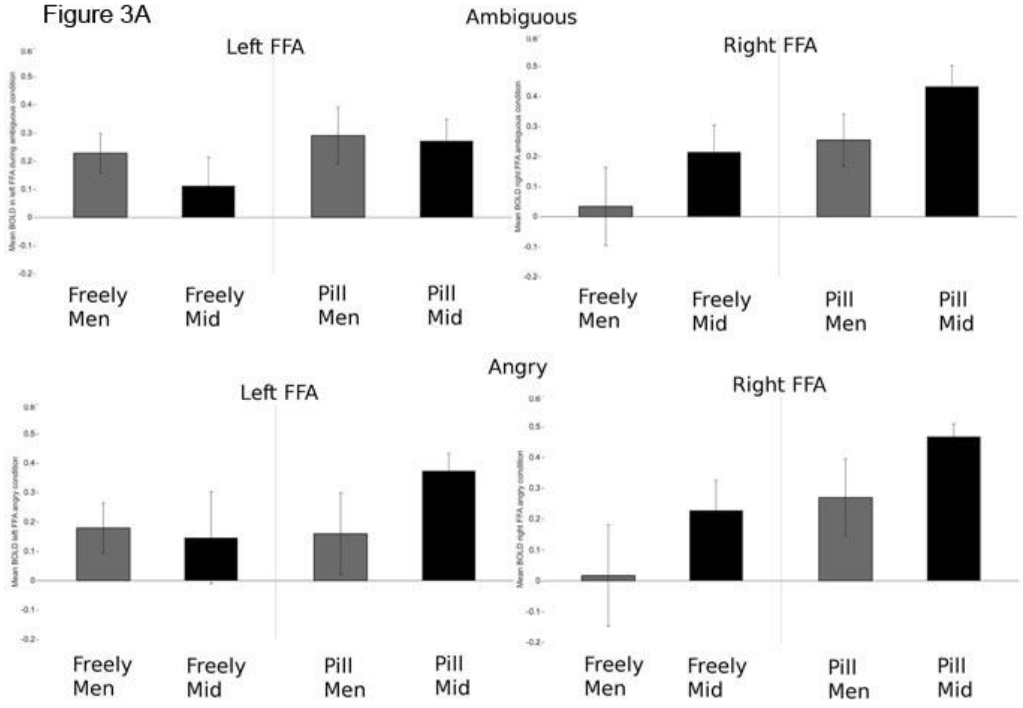
5.3.2.1 Functional MRI: voxelwise analysis

The pill vs. freely cycling contrast showed one significant cluster in the right caudate region (z=3.71; x=12, y=2, z=16) in the angry condition and no significant clusters in the ambiguous condition. The freely vs. pill contrast showed one significant cluster in the right thalamus region (z=3.87; x=6, y=-12, z=14) in the angry condition and no significant clusters in the ambiguous condition.

5.3.2.2 Functional MRI: ROI analysis of the FFA response

Mean BOLD response in the left FFA was higher in the Pill group (vs. No Pill group) during the ambiguous condition ($t_{(108)}$=2.57, p=0.012, Cohen's d=0.49) and remained significant after correcting for multiple comparisons. Mean BOLD response in the left FFA during angry condition and right FFA during both ambiguous and angry condition did not show any significant difference when comparing adolescent girls in the pill vs. no pill group (left FFA angry: $t_{(108)}$=1.05, p=0.3; right FFA ambiguous: $t_{(108)}$=1.46, p=0.15; right FFA angry: $t_{(108)}$=0.41, p=0.68). These results are illustrated in Figure 3B.

A 2-way repeated-measures ANOVA examining the effect of Pill, Condition, and Pill*Condition interaction on BOLD response in the left FFA confirmed the main effect of Pill (F=4.0, p=0.048), and showed no effect of Condition (F=0.88, p=0.35) or Pill*Condition interaction (F=1.59, p=0.21). A 2-way repeated-measures ANOVA examining the effect of Pill, Condition, and their interaction on BOLD response in the right FFA showed no significant results (Pill: F=0.97, p=0.33; Condition: F=0.001, p=0.98; Pill*Condition: F=0.85, p=0.36). Adding Laterality to the whole model and testing it with a 3-way ANOVA confirmed the effect of Pill (F=7.33, p=0.007) and showed no effect of Laterality (F=2.5, p=0.11), Condition (F=0.13, p=0.72), or any related interaction (Pill*Laterality: F=0.7, p=0.4; Pill*Condition: F=0.17, p=0.68, Pill*Laterality*Condition: F=0.04, p=0.84) on the mean BOLD response in FFA.



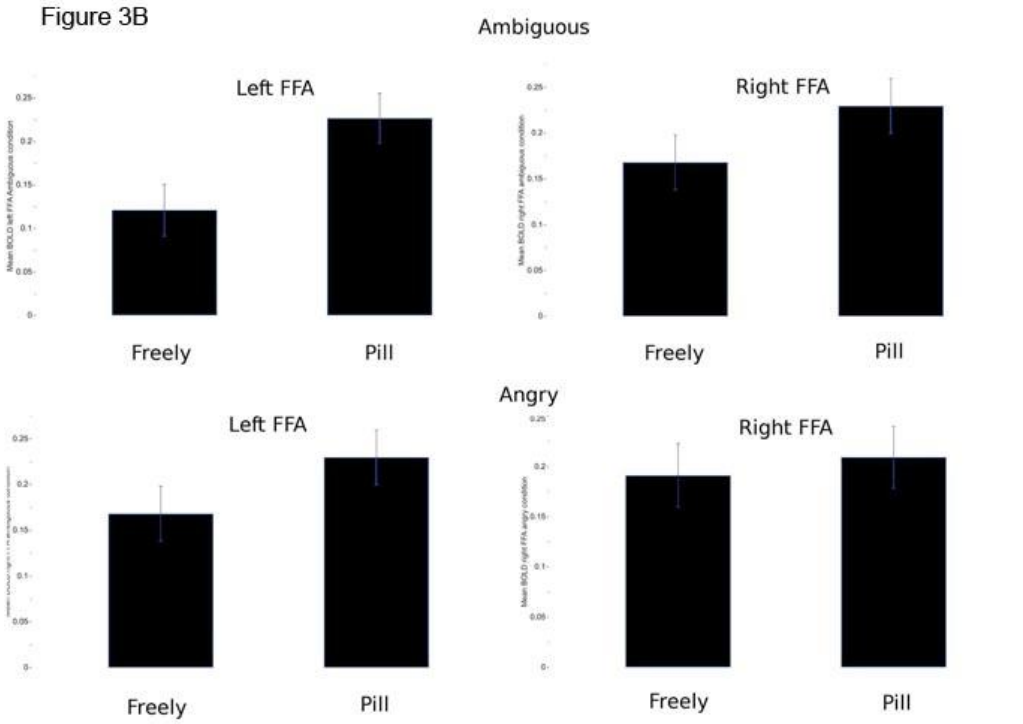


Figure 3A: Experiment I – Effect of pill and cycle phase on BOLD response in FFA in young women

Figure 3B: Experiment II – Effect of pill on BOLD response in FFA in adolescent females

5.3.2.3 Functional MRI: ROI analysis of the amygdala response

Upon reviewer's suggestion we have conducted ROI analysis of the amygdala response also in Experiment II and we report these results in Appendix 8 now.

5.3.2.4 Personality: NEO-FFI

No significant differences were found in the personality profiles of OC vs. freely cycling adolescent girls: neuroticism (OC: M=23.51, SD=8.00; freely cycling: M=25.25, SD=6.84; $t_{(108)}$=-1.23, p=0.22), extraversion (OC: M=30.311, SD=5.27; freely cycling: M=28.96, SD=5.58; $t_{(108)}$=1.3, p=0.2), openness (OC: M=25.58, SD=5.48; freely cycling: M=27.16, SD=5.63; $t_{(108)}$=-1.49, p=0.14), agreeableness (OC: M=29.29, SD=5.88; freely cycling: M=29.4, SD=4.79; $t_{(108)}$=-0.11, p=0.92), and conscientiousness ((OC: M=28.38, SD=7.03; freely cycling: M=28, SD=6.96; $t_{(108)}$=0.29, p=0.78).

5.4 EXPERIMENT III: SCANNING EYE MOVEMENTS

5.4.1 Methods and materials

In order to assess whether the above group differences in the FFA response to faces might be explained by a different pattern of eye movements while scanning the face, we conducted a behavioral eye-tracking study in another sample of women. Morris et al. (2007) experimentally manipulated scanpath during face viewing and showed

that atypical scanpaths (i.e. less fixations to the eye region) are associated with lower BOLD response in the FFA.

5.4.1.1 Participants

Twenty healthy women between the ages 18 and 29 years (*M*=24.5 years, *SD*=3.07) were recruited from the Baycrest volunteer database and Toronto area: 10 freely cycling and 10 taking OC. Participants' consent was obtained according to the Declaration of Helsinki and approved by the Ethics Committee at Baycrest. The five types of OC used (Alesse, Cyclen, Diane, Marvelon, Yasmine) have 20–35 µg of Ethinyl Estradiol and different progestin substances (200 µg of Cyprosterone in Diane, 150 µg of Desogestrel in Marvelon, 300 µg of Drospirenone in Yasmine, 100 µg of Levonogestrel in Alesse, 250 µg of Norgestimate in Cyclen). Duration of OC ranged from 3 to 180 months (*M*=47.9 months, *SD*=56.25). There was no significant relationship between OC duration and age (*r*=0.35, *p*=0.33). Participants were tested during the 2^{nd}, 3^{rd}, or 4^{th} day after the end of their period. This timing was chosen to maximize the difference in estrogen levels between the pill and freely cycling women; while women taking OC were already receiving artificial estrogens, freely cycling women had still very low levels of endogenous estrogens

5.4.1.2 Procedure

An eye-tracking system (EyeLink II) was used to record eye movements of the participants while they were watching the angry and

ambiguous face videoclips used in Experiment I. Participants were sitting 91 cm from the monitor and were viewing 8-cm long faces under a visual angle of 5 degrees; this corresponds to the angular size employed in Experiment I. Chin rest was used to minimize head movements. The eye-tracker was calibrated before each 18-s block of videoclips.

5.4.2 Results

A mixed model exploring the effects of OC use (OC, freely cycling), Type of Face (ambiguous, angry), and Face Area viewed (interest areas: eyes, nose and mouth) on dwell time explained 53% of variability in the data (Adj R^2=0.53) and showed main effects of Face Type ($F_{(1,18)}$=90.93, p<0.0001), Interest Area ($F_{(2,36)}$=29.76, p<0.0001), and an interaction between the two ($F_{(2,36)}$=42.28, p<0.0001); the dwell times were longer while viewing ambiguous (vs. angry) faces and when looking at the eyes (vs. nose vs. mouth), with the latter effect more pronounced while viewing ambiguous faces. There was no main effect of Pill ($F_{(1,18)}$=0.79, p=0.39) or any pill-related interactions (Pill*Face Type: $F_{(2,18)}$=0.02, p=0.88; Pill*Interest area: $F_{(2,36)}$=3.1, p=0.06; and Pill*Face Type*Interest Area: $F_{(2,36)}$=1.14, p=0.33).

5.5 DISCUSSION

Results of the voxelwise analysis showed that women taking OC compared with freely cycling women, as well as women during mid-cycle compared with menstruating women, had stronger BOLD

response to faces in the right FFA, and several other brain regions (e.g. IFG, STS, middle temporal gyrus). This finding is consistent with those of Dietrich et al. (2001) who scanned women performing a verbal task and mental rotation tasks at different phases of menstrual cycle and showed that female sex hormones had a profound effect on the magnitude, but not lateralization or localization, of brain responses associated with these tasks. The fact that these pill/phase-related effects reported in our study were most pronounced in the right FFA is also consistent with our previous observation, made in a large group of adolescents using the same face paradigm, of the largest sex difference in the brain response to faces (female > male) being present the right FFA (Tahmasebi et al., 2012).

Focusing on the FFA, we also found that the mean BOLD response in this cortical region increased as a function of OC use duration; women who used OC longer had a stronger BOLD response in the left FFA while viewing both ambiguous and angry faces, and in the right FFA while viewing angry faces. Even though these effects of OC duration did not survive correction for multiple comparisons, they are suggestive of a dose-response relationship between the OC use and the brain response to faces.

As expected, plasma levels of estradiol were significantly higher in the freely cycling women compared with women taking OC (but note the high inter-individual variations in hormone levels). It should be noted, however, that this observation concerns endogenous sex hormones

only (see below) and, as such, can be potentially misleading with regards to the interpretation of our findings. Standard assays are designed to measure these naturally occurring (endogenous) forms of estradiol. Contraceptive steroids, however, show only a limited cross-reactivity with the antiserum used in the standard assays (Hampson, & Young, 2007). Therefore blood samples processed with the standard assays do not reflect the overall levels of female sex-steroids (Hampson, & Young, 2007). But artificial sex steroids (ethinyl estradiol) contained in OC are biologically active: “ethinyl estradiol binds to the estrogen receptor complex and enters the nucleus activating DNA transcription of genes involved in estrogenic cellular responses” (Medical Dictionary, 2011). Oral contraception provides constant levels of estrogen and progestin in the blood and thus suppresses the pulsatile release of FSH and LH from the anterior pituitary; during the pill-active phase, women taking OC are exposed daily to a high influx of ethinyl estradiol. The daily peak plasma concentrations of ethinyl estradiol in women taking OCs with 30 µg of ethinyl estradiol range from 125 to 168 pg/mL (Van der Heuvel et al., 2005), corresponding to 459 to 616 pmol/L (Society for Biomedical Diabetes Research, 2010). Since participants in Experiments I and II were taking OC with 30 µg and between 20 and 35 µg of ethinyl estradiol, respectively, their daily peak plasma levels of ethinyl estradiol were likely to be in a similar range as reported by Van der Heuvel et al. (2005) and measured in freely cycling women in the current study (estradiol) during mid-cycle (597.68 pmol/L).

An indirect evidence of higher estrogen levels in OC women comes from studies exploring the effects of estrogen on tasks that are known to show a female “advantage”, such as episodic (Herlitz & Rehnman, 2008) or verbal (Kimura, 1999) memory, or female “disadvantage”, such as spatial skills (Geary & DeSoto, 2001). Rosenberg (2002) showed that women taking OC reached better verbal scores but worse spatial scores compared with freely cycling women. Worse performance on visuospatial task in OC vs. freely cycling women was also reported by Wharton et al. (2008). Mordecai et al. (2008) reported that OC users showed enhanced verbal memory during the active pill phase, while freely cycling women did not show any difference. We have shown clear sex differences using the face task employed here: female adolescents have a stronger FFA response to ambiguous faces than male adolescents (Tahmasebi et al., 2012). Thus, we observe a similar directionality in the FFA response to faces across sex (female > male), cycle (mid-cycle > menstruation) and OC (pill > no pill). Nonetheless, whether estradiol is the common element binding together these three phenomena remains to be tested in future studies.

Interpretation of the observed correlations between (endogenous) estradiol and BOLD in right FFA is challenging. First of all, the fact that these hormone-brain relationships hold only when the estrogen levels are very low (menstruation) suggests a fast saturation of the hormonal effects. Second, the opposite direction of this relationship in the two groups might reflect either the different state (“plasticity” – see below) of

the system in the long-term users of OC and/or the reciprocal relationship between endogenous and synthetic hormones in OC women. The latter could be related, for example, to the variation in the rate of EE metabolism: slow metabolizers of EE may still have relatively high levels of EE – and therefore low levels of (endogenous) E – in their system during menstruation and, as such, the observed inverse relationship between E and BOLD may, in fact, reflect positive relationship between EE (not measured) and BOLD.

While the voxelwise results of Experiment II did not replicate findings from Experiment I, results of the ROI analysis of experiment II showed an enhanced BOLD response in FFA in OC compared with freely cycling adolescent girls, thus replicating results obtained in Experiment I in a sample of adolescent girls. Note that we designed Experiment II as a replication of the FFA effects observed in Experiment I; voxelwise analysis was added only for completeness of the reporting. The effect of pill on FFA BOLD response was in the medium range for both samples, but still slightly higher in the adults compared with adolescents. The fact that the group differences in the FFA response to faces are smaller in the adolescent sample, as compared with the young women, is consistent with the overall shorter duration of the OC use in the adolescent sample. Smaller effect size in the sample of adolescents might be also related to the fact that the adolescent girls were not tested during a particular phase of their cycle, which could have added noise to the pill effect. As shown in Figure 3B, both right and left FFA showed the same direction for the effect of pill (pill>freely)

on BOLD response and a 3-way ANOVA examining the effects of pill, condition, and laterality on the FFA BOLD response showed an effect of pill but no effect of laterality or any laterality-related interaction.

Results from Experiment III showed that women taking OC did not differ from freely cycling women in their pattern of exploratory eye-movements while viewing the same videoclips of faces employed in Experiment I. Although Dalton et al. (2005) showed that persons with autism who made fewer fixations on the eyes of a displayed face had a lower BOLD response in FFA than those who made more fixations on the eyes, and Morris et al. (2007) manipulated scanpaths within healthy individuals and found that "atypical" scanpaths of faces were associated with lower BOLD response in FFA, both authors agree that under free-viewing condition healthy participants did not show any correlation between scanpath and FFA response. If not the pattern of eye movements, what else might underlie the observed group differences in the FFA response?

It is possible that women taking OC pay closer attention to the cues carried by the face when processing, for example, the biologically significant regions of the face, such as the person's eyes and their direction (gaze). It is known, for example, that attention to visual cues increases activity in extrastriate visual cortex and anterior temporal region (Lane et al. 1999). Kastner et al. (1999) demonstrated that directing attention to a particular location, and expecting the occurrence of visual stimuli at that location, increases BOLD response of human

visual cortex even in the absence of visual stimulation. Palermo & Rhodes (2007) supported these results and showed that the neural responses in face-selective cortex are larger for attended compared with ignored faces. Petrovic et al. (2008) showed that during evaluation of conditioned faces oxytocin modulated BOLD response in the fusiform gyrus and amygdala. Given the modulatory role of estrogens on oxytocin (McCarthy, 1995; Stock et al., 1994; Silber et al., 1987), future research might consider the possible effect of oxytocin on attention and the possibility that estrogen effects on brain response to faces might be mediated by oxytocin.

We do not know whether the effect of OC on brain response to faces has any behavioral consequences. It is possible that women taking OC might demonstrate a better detection and/or recognition of faces in tasks where sex differences between males and females have been described (McBain et al., 2009, Hall & Matsumoto, 2004, Hampson et al., 2006, Kirouac, & Dore, 1985, Rahman et al., 2004; Montagne et al., 2005). Since these studies report small to medium effect size for differences between males and females, it is very likely that the effects of OC would be only small and a larger sample size would be necessary to detect them.

While replicating the effect of pill on BOLD response in FFA in two independent samples (Experiment I and II) is strength of this study, using a third sample for collecting purely behavioral data is a limitation. Conducting a multimodal study that would record scanpaths during the

fMRI task and had a bigger sample size would be ideal. Nonetheless, we used the same recruitment criteria (women, 18-29 years old, of white Caucasian background) and the same stimuli (Grosbras & Paus, 2006) in Experiment I and III.

Use of OCs exposes women to exogenous estrogen and progesterone, suppresses the release of FSH and LH and lowers levels of endogenous estrogens and progesterones. Use of OCs also lowers levels of testosterone (Graham et al., 2007; Hietala et al., 2007). In the current study, both the pill vs. no pill and the mid-cycle vs. menstruation contrasts showed a higher BOLD response in the right FFA, suggesting a similarity in the modulation of the brain response to faces in this region between the two hormonal states. But the exact mechanisms underlying such a similarity are unknown and may include, for example, OC-related variations in other hormones (e.g. low levels of testosterone) or long-term effects of OC at the level of the brain (see below).

Future studies may explore, for example, the effect of progestin-only pills on brain response to faces; this would allow us to ascertain whether the observed effects are related to the progestin part of the combined oral contraceptives used by women included in the present study or whether they might be related to the stimulating effects of estradiol. It would be also helpful to measure levels of androgens. But the exact mechanisms underlying effects of sex hormones are best addressed in experimental models. In this context, it is of interest to

note the work of Follesa et al. (2001) who showed that long-term exposure to OC decreases levels of progesterone and its metabolite allopregnanolone in the rat brain, and in turn, increases expression of GABAa subunits in the cerebral cortex. Similarly, Smith (1994) reviewed several mechanisms by which estradiol might amplify neuronal excitability, including those involving excitatory (e.g., synthesis, degradation and release of glutamate) and inhibitory (reduction of GABA) neurotransmission. It is unclear, however, how such effects combine over time and whether the net outcome of the long-term use of OC is an increase or decrease in cortical excitability. The positive relationship between the duration of OC use and the FFA response to faces observed in our study suggests that the net effect of OC might be that of a higher excitability; this hypothesis can be explored, for example, in studies of cortical excitability with paired-pulse transcranial magnetic stimulation (Wassermann et al., 2008).

5.6 CONCLUSION

This study examined the effects of female sex hormones on BOLD response to faces. Women taking oral contraception (vs. freely cycling women) and women during mid-cycle (vs. menstruation), were found to have a stronger BOLD response in the right fusiform face area (and several other cortical regions). Stronger BOLD response to faces in the fusiform face area was also found when comparing OC and freely cycling adolescent girls. The behavioral study suggests that these group differences are not related to the pattern of exploratory eye

movements while viewing faces. The mean BOLD response in FFA also increased as a function of OC use duration, supporting the possibility of a long-term plasticity-like adaptation related to the use of OC.

Next chapter will summarize main findings of this thesis and discuss them in the context of organizational and activational hypothesis, other domains of cognition, and evolutionary perspectives.

CHAPTER 6

CHAPTER 6

Discussion

This thesis studied the role of sex hormones as a potential mechanism that contributes to the development of sex differences in craniofacial features and face processing. Craniofacial features were quantified using MR images of the head and the relationships between craniofacial features and exposure to prenatal and pubertal androgens were studied using discordant-sex twin design and levels of bioavailable testosterone from serum, respectively. We showed that development of craniofacial features was associated with exposure to androgens during both pre-natal and post-natal period. Craniofacial features characteristic for exposure to prenatal androgens were measured also in an independent dataset of adolescent females and relationship between the index of prenatal androgens in the face and brain size was explored.

Face processing was studied using functional magnetic resonance imaging (fMRI) and tracking of eye movements. We focused on the role of postnatal sex hormones and used phase of menstrual cycle and use of oral contraception to predict brain response to faces and eye-movements scanning the face. We will now review the main

findings of each chapter and highlight the main contributions of the thesis to the current literature.

6.1 MAIN FINDINGS

In Chapter 2, we used faces reconstructed from MR images to describe sex differences in face shape during adolescence. We showed that males had wider face, shorter nose, bigger jaw, and narrower eyes than females. These typicaly male facial features emerged as a function of age-adjusted bioavailable testosterone in both males and females. We presented these MRI-reconstructed faces to (adult) raters and asked them to provide their impression about the sex of the face. Raters correctly classified faces of older male adolescents but performed at chance when classifying the sex of 12-year old male faces and the faces of all female adolescents (irrespective of their age). Still, there were some female faces that were consistently classified correctly, and those that were consistently classified incorrectly (as belonging to males) by most raters. These correctly and incorrectly classified female faces did not differ in age but did differ in the presence of testosterone-related facial features typical for males; that is female faces mis-classified as males had higher levels of these male-type facial features. Females incorrectly classified as males had also more body fat than the correctly classified females. We investigated to what extent it is the fat and to what extent it is the actual skull shape that contributes to the development of sex differences in the face. We tried to answer this question initially by adjusting facial features for total body

fat and re-running our analyses. An exact answer for this quesiton, however, was not possible when using the MRI-reconstructed faces (based on facial tissue), and thus we decided to study features of the skull, as a relatively ‘fat-free’ metric.

We placed skull landmarks on head MR images and explored the relationships between skull shape, face shape, and sex judgments about the face (Chapter 3). Skull shape mediated the relationship between face shape and sex judgments about male but not female faces. While body fat had a slight positive effect on correct sex recognition of male faces, it had a negative effect on correct sex recognition of female faces. The mediation of the relationship between face shape and correct sex recognition by skull was seen in females only after adjusting the female face shape for body fat.

In Chapter 4 (Study 1), we explored whether prenatal androgens might influence the development of facial features. We used a discordant-sex twin design to evaluate the effect of co-twin’s sex on face shape. Since the intra-uterine presence of a co-twin brother (vs. co-twin sister) increases exposure to prenatal androgens in the other twin, we hypothesised that faces of individuals with a co-twin brother will differ from faces of individuals with a co-twin sister. In order to avoid the potential effect of fat on face shape, we placed skull landmarks on the head MR images of the twins and studied the relationship between twin group and shape of the skull. Females with a co-twin sister differed from all the other twin groups that were exposed to prenatal androgens

(females with co-twin brother, males with co-twin sister, males with co-twin brother): they had shorter distance between the inside corners of the eye sockets, larger distance between left and right third molars of the lower jaw, and larger distance between the left third molar and lower front-teeth, right third molar and lower front-teeth, left third molar and tip of the chin, and right third molar and tip of the chin. These features were captured by a third principal component (PC3). The size of this effect was large. We did not find any difference in skull shape of males with a female (vs. male) co-twin. We concluded that prenatal androgens did leave their signature in the face and that this effect appears at a certain level of prenatal androgens but does not follow a simple (linear) dose response.

Since direct measures of prenatal androgens are often not available and access to data from individuals exposed to higher vs. lower levels of prenatal androgens (e.g. twin studies or individuals with complete androgen insensivity syndrome or congenital adrenal hyperplasia) is sparce, we aimed to test whether the signature of prenatal androgens in the face, identified in the twin study, might be used as a proxy of exposure to prenatal androgens – similarly to digit ratio. Having such an indirect measure of prenatal androgens would enable us to estimate the exposure to prenatal androgens in all individuals for whom we collected head MRI data (n=2,000 in *Imagen*, n=1,024 in *SYS*) and possibly study the interaction between prenatal (facial signature) and postnatal sex hormones (serum testosterone, oral contraception use) and their possible effect on cognition or disease risk.

We aimed, accordingly, to find the signature of prenatal androgens in the face in an independent dataset and then tested its relationship with a variable that is known to be modified by levels of prenatal androgens (Chapter 4, Study 2). We placed the 19 skull landmarks, used in the twin study (Chapter 4, Study 1), on MRIs of SYS participants and calculated similarity between craniofacial features of each SYS participant and the PC3 model describing signature of prenatal androgens in the face. Since brain size is associated with exposure to prenatal androgens (e.g. Pepper et al., 2009), we tested the relationship between facial signature and brain size in adolescent females from SYS. Our results supported the existence of signature of prenatal androgens in the face because we showed that females with more (vs. less) prenatal androgens-related craniofacial features had also slightly larger brain size. The facial signature could explain up to 8% of variance in brain size of females from SYS.

The fifth chapter studied effects of sex hormones on face processing. Literature showed that females performed better in detection of faces (McBain et al., 2009), recognition of sex of the face (Chapter 2), and recognition of emotion displayed by the face (e.g. Hampson et al., 2006). Females had also higher brain response to faces than males (Tahmasebi et al., 2012). Further research showed that ovarian sex hormones did influence performance on emotion recognition tasks (Derntl et al, 2008; Pearson, & Lewis, 2005). We tested whether the ovarian sex hormones might also modulate brain response to faces (Chapter 5, Experiment I). Phase of menstrual cycle

and use of oral contraception served as predictors of BOLD response in adult female participants observing angry and ambiguous face videoclips. We showed that females in mid-cycle (vs. menstruation) and females taking oral contraception (vs. freely cycling) showed higher BOLD response in fusiform face area. We also showed that the BOLD response in FFA varried as a function of duration of oral contraception use but these results should be interpreted with caution due to small sample size. The main effect of oral contraception on FFA brain response was replicated in an independent sample of adolescent females (Chapter 5, Experiment II).

An eye-tracking experiment was designed to follow these effects of oral contraception use on brain response to faces (Chapter 5, Experiment III). Participants were presented with the same face videoclips as used in Chapter 5, Experiment I and II and the dwell time and number of fixations in regions of the face (eyes, nose, mouth, rest of the face) were calculated. No differences between females taking oral contraception and freely cycling females were found which suggested that the relationship between ovarian sex hormones and brain response to faces could be related to covert attention.

Overall we conclude that sex hormones do, indeed, influence both the development of face shape and the way in which faces are processed. The findings here suggest that the effects of these hormones appear to contribute to the development of sex differences in face development and face processing. The thesis has also

demonstrated, and replicated, the effect of oral contraception on brain response to faces. Considering the wide use of oral contraception worldwide (Mosher et al., 2004), further research about the effects of oral contraception use on brain and behavior are imperative.

Methodologically, a strength of the thesis is the novel use of head MR images and MRI-face reconstruction to study craniofacial features. This method resulted in the identification of a signature of prenatal androgens in the face that can be used as an indirect index of exposure to prenatal androgens in future studies, as an additional or alternative to digit ratio.

Chapters 2-5 discussed the main findings in the context of current literature. In the current chapter, these findings will be discussed in the context of (1) organizational and activational hypothesis/ the alternative extended critical window hypothesis, (2) other domains of cognition, (3) evolutionary perspectives. The limitations of the research presented in the thesis will be discussed and possible topics and questions for future research will be explored.

6.2 ORGANIZATIONAL AND ACTIVATIONAL EFFECTS OF SEX HORMONES ON FACE AND THE BRAIN

According to the organizational and activational hypothesis (Phoenix et al., 1959), the effects of prenatal sex hormones are organizational and permanent while the effects of postnatal sex hormones are activational, acute and reversible. While true

experimental testing of the organizational and activational hypothesis is possible only in experimental animals, and studying its mechanisms in humans is limited by the lack of access to prenatal androgens, the principles of organizational and activational hypothesis are an important basis of research questions about the role of sex hormones in face development and face processing. Moreover, in this thesis, the lack of available prenatal androgen data inspired the development of a facial signature that could be used as an indicator of prenatal androgens.

Currently, digit ratio is the only readily available - albeit indirect - index of prenatal androgens. Numerous studies, however, questioned its reliability (e.g. Buck et al., 2003; Berenbaum et al., 2009; Medland et al., 2010; Lilley et al., 2010). Many authors were able to show an association with digit ratio only when using right but not left hand (or vice versa; reviewed in McIntyre, 2006). Having another readily available and, possibly, more reliable index of prenatal androgens would thus be useful.

We used a discordant-sex twin design and MR images of the head to identify a signature of prenatal androgens in the face (Chapter 4, Study 1). Since the effect of co-twin's sex on face shape was large (Cohen's d = 0.76), we assessed this facial signature in an independent dataset, showed its relationship with brain size, a known correlate of prenatal androgens (Peper et al., 2009), and suggested that this facial signature might be used as an indirect index of prenatal androgens (Chapter 4, Study 2). Facial signature was able to predict similar portion

of variance in brain size (cca 2%) as discordant-sex twin design used in Peper et al (2009). Future research is necessary to further verify the relationship between facial signature and levels of prenatal androgens measured directly from amniotic fluid and test reliability of the facial signature as an estimate prenatal androgens in other studies.

It is possible that we might fine-tune the facial signature by focusing on the chin and jaw only (since these reflect most of the PC3-related features) and thus reducing possible noise that might be coming from the rest of the face. Alternatively, it is possible that the effects of prenatal androgens on face shape might be more pronounced in individuals with a particular gene variant, as suggested, for example, for gene SMOC 1 in the context of prenatal androgens and digit ratio (Lawrence-Owen et al., 2012). We speculate, that individuals with particular gene variant might show a stronger relationship between prenatal androgens and the face shape and thus allow better prediction of prenatal androgens-related effects (as indexed by the facial signature) on cognition or disease risk.

Further research is necessary to explore the organizational and activational role of sex hormones on brain function. So far, we have demonstrated the effects of menstrual cycle and oral contraception use on face processing and it seems that the effects of menstrual cycle support the Phoenix et al. (1959) finding about the acute and reversible activational effects of postnatal sex hormones. It is not clear, however, whether the effect of duration of oral contraception use on brain

response to faces in adult women is also reversible. A longitudinal research would need to clarify whether the plasticity-like effects of OC duration disappear after discontinuation of OC or whether they might leave any permanent (organizational) effects.

Future research might also benefit from the exploration of possible interactions between prenatal and postnatal sex hormones on brain response to faces. For example, would females with lower (vs. higher) exposure to prenatal androgens in utero have larger brain response to faces and/or better emotion recognition skills? Also, would females with higher (vs. lower) levels of androgens *in utero* show different effects of menstrual cycle and oral contraception use on brain response to faces? We predict that females who take oral contraception (vs. freely cycling) and were exposed to lower (vs. higher) levels of androgens in utero would have the largest brain response to faces and were most , accurate in emotion recognition. A dataset including (i) measures of prenatal androgens from amniotic fluid data from dizygotic twins and (ii) information about postnatal sex hormones, and (iii) fMRI and behavioral data about face perception would be ideal to answer these questions.

6.3 SEX HORMONES AS THE UNDERLYING MECHANISM OF SEX DIFFERENCES IN FACE PROCESSING AS WELL AS OTHER DOMAINS OF COGNITION

Our findings on the female (vs. male) advantage in correct sex recognition (Chapter 2) are consistent with the literature on sex differences in face detection: females are more accurate than males in

(i) the detection of upright faces, but not trees (McBain et al., 2009); and (ii) in recognition of emotions in the face (Hampson et al., 2006; Hall & Matsumoto, 2004).

Our findings on the effects of menstrual cycle and oral contraception on brain response to faces (Chapter 5) suggest that ovarian hormones might contribute to the fine-tuned ability of face perception in females *. These findings are consistent with (i) the simple sex difference in brain response to faces, where females (vs. males) had stronger BOLD response to faces in almost all regions of the face processing network except for amygdala and rhinal sulcus (Tahmasebi et al., 2012), and (ii) the effect of menstrual cycle on facial emotion recognition, where better fear recognition skills were observed during pre-ovulatory compared with menstruation phase (Pearson & Lewis, 2005).

While literature suggests that FFA responds not only to faces but reflects expertise in any type of objects in general – e.g. cars or birds (e.g. McGugin et al., 2012; Gauthier et al., 2000), we chose to study brain response to faces in particular because that was an area where consistent sex differences in behavior have been described in the past (e.g. Hoffmann et al., 2010) and where we would expect to find, based on the evolutionary hypotheses (e.g. Hampson et al., 2006; Penton-Voak, & Perrett (2000), the female advantage and potential role of sex hormones.

* Please note that the purpose of our research was to study the effects of sex hormones on face perception. We did not mean to parse out the face effects from the emotion. We do not consider our ambiguous face condition as an appropriate control for the angry condition for two reasons: First, Tahmasebi et al. (2012) observed more BOLD response during the ambiguous condition than the angry condition. Second, also shown in Tahmasebi et al. (2012), and more relevant to the present paper, the sex differences in many brain regions (including the FFA) were more pronounced during the ambiguous condition than the angry condition. This may be due to a few factors present in the ambiguous condition, including but not limited to: the lack of repetition in the facial movements, increased difficulty in interpreting the valence of the face, etc.Notably, we consider the comparison of our dynamic social videos with videos of non-social stimuli biologically appropriate, and consistent with past research. The contrast of faces vs. moving circles is similar to the contrasts employed in Hariri's well accepted Face task design (Hariri et al., 2002), which contrasts responses during perceptual matching of angry and fearful faces versus perceptual matching of shapes. Investigation of the two face conditions might be rather thought of as a within study replication of the sex hormones effects. Future research might want to study the effects of sex hormones on emotion.

It may be that face perception is a further example of consistent sex differences in cognitions that can, in part, be explained by sex hormones. Sex differences have been described in spatial and verbal skills: males outperformed females in spatial abilities across all ages (Voyer et al., 1995). Males also demonstrated better performance on episodic memory task involving spatial processing (Lewin et al., 2001). In contrast, females outperformed males in verbal skills, particularly in speech production (Hyde & Linn, 1988), verbal association (Hines, 1990), and verbal episodic memory (Herlitz et al., 1997). The role of ovarian sex hormones as the underlying mechanism of these sex differences has been suggested in both spatial (e.g. Rosenerg & Park, 2002) and verbal skills (e.g. Maki et al., 2002). Fluctuations of sex hormones due to menstrual cycle were associated with performance in verbal and spatial tasks (Halpern, & Tan, 2001) and altered sex

hormone levels due to oral contraception improved verbal memory (Mordecai et al., 2008).

6.4 EVOLUTIONARY PERSPECTIVES ON SEX DIFFERENCES IN FACE PROCESSING

It is possible that some of these sex differences in cognition might have developed evolutionarily. Hampson et al. (2006) suggested that female advantage in recognition of emotions in the face might be influenced by evolution and related to women's responsibility for child-rearing. They tested two variants of the child-rearing hypothesis: (1) attachment promotion hypothesis, which says that infants whose mothers were responsive to their nonverbal signals (e.g. cry, smile) have more secure attachment (Ainsworth, 1979; Hall et al., 1986), optimal long-term health, immune function, and social outcomes (Goldberg, 2000) than infants whose mothers were not as responsive; and (2) fitness-threat hypothesis, which says that infants were more likely to survive if their mothers were responsive to their negative emotions, signals of pain and potential threat. While attachment promotion hypothesis suggests that females would have advantage in recognition of any type of emotional expression, fitness-threat hypothesis suggests that females would have an advantage in recognition of negative emotions (Hampson et al., 2006).

Hampson et al. (2006) showed that females outperformed males in recognition of both positive and negative emotions, but the effect size of the sex differences was larger in the negative compared with positive

emotion recognition. Both the attachment promotion hypothesis as well as the fitness-threat hypothesis were thus supported, showing that social information processing is more fine tuned in females. Their study also pointed out that the female advantage in recognition of emotions in the face was not related to previous childcare experience or a generally higher perceptual speed (Hampson et al., 2006).

It is possible that oxytocin, a neuropeptide involved in social behavior (Ellenbogen et al., 2012; Groppe et al., 2013), might provide a biological explanation of the child-rearing hypothesis. Domes et al (2012) showed that intranasal oxytocin increased covert attention to happy faces, suggesting that oxytocin modulates early attentional processes and promotes prosocial behavior. Lischke et al. (2012) reported a relationship between levels of oxytocin and performance on emotion recognition task involving dynamic face stimuli and showed that it was independent of overt visual attention measured by eye-tracking. Since females tend to have higher levels of oxytocin than males (Kramer et al., 2004; Zingg, 2002), they might pay more attention to faces and thus be able to recognize correctly even slight changes in facial expression (Hoffmann et al., 2010). Levels of oxytocin were found to be higher also in women taking oral contraception compared with freely cycling women (Stock et al., 1994; Silber et al., 1987). Future research should clarify whether oxytocin might potentially mediate the relationship between ovarian sex hormones and performance on emotion recognition task reported in e.g. Derntl et al (2008) or Pearson, & Lewis (2005).

Penton-Voak and Perrett (2000) suggested that sexual selection might be another evolutionary mechanism that might help to explain the female advantage in face perception. They reported that women during folicular vs. luteal or menstrual phase of menstrual cycle showed enhanced sensitivity to reproductively relevant stimuli and prefered more masculinised faces. Since follicular phase of menstrual cycle precedes ovulation, higher sensitivity to facial cues during this period might be important for mating. Our findings of higher brain response to faces during mid-cycle vs. menstruation (Chapter 5) are consistent with this hypothesis.

6.5 PRACTICAL SIGNIFICANCE OF THE RESULTS

Our findings may have practical significance in three main areas: (1) research that recruits women taking OC, (2) research that would like to develop an easily accessible indicator of hyperandrogenism in women, (3) research that is interested in the effects of prenatal androgens but did not assess levels of prenatal androgens directly.

First, considering the number of women using oral contraception worldwide (100 million women according to Mosher et al., 2004) and the effects of OC duration on brain response to faces (Mareckova et al., 2012), more research about the effects of OC on the brain is imperative. Most studies of the effects of sex hormones still exclude women using OC. Studies that try to study effects of OC often have insufficient power to consider (1) the different composition of OCs available on market as well as (2) the fact that women often discontinue OC use during certain

periods of life or decide to change their OC brand. More research with adequate sample size to consider these conditions is needed. The current state of research indicates that OC use and its duration should be considered as a covariate in other studies that do not focus on the effects of OC in particular.

Second, considering that faces of adolescent females with high levels of bioavailable testosterone were often misclassified as those of males, and their faces included many male-like features (e.g. wider face, shorter nose, bigger jaw, and narrow eyes), the presence of these features in the female face might be used as an easily accessible index of hyperandrogenism and related health issues which might not be otherwise noticed. Having such an easily accessible index would likely trigger early diagnosis and prevention of potential health issues (e.g. polycystic ovary syndrome) that might appear later in life. In addition, the current research on sex-specific face and skull features provides biomarkers that might be used in facial reconstruction after injury or gender reassignments.

Third, since face shape differs between females with same-sex vs. opposite-sex twins, it is likely that prenatal androgens might have left a signature in the face. Such a signature might be used as an alternative to digit ratio, the widely used but often criticised index of prenatal exposure to androgens, in studies that are interested in the effects of prenatal androgens, collected MR images, but did not measure prenatal androgens directly.

6.6 LIMITATIONS AND FUTURE RESEARCH

There are four main issues that limit our conclusions and should be resolved in future research (see bellow). In addition, we realize that most of our studies were focused on females and that it might be interesting, for example, to explore the potential role of bioavailable testosterone on brain response to faces in males. We also realize that the online survey used for the sex judgments task (described in Chapter 2 and Chapter 3) did not allow verification of participants' identity and controlled environment and that a laboratory setting would be more appropriate for this task. The lack of longitudinal data that would allow us to clarify the reversibility/permanence of the effects of sex hormones during adolescence

It is difficult to determine which of our findings are only acute and reversible and which are permanent. Longitudinal research would be necessary in order to decide whether sex hormones might have organizational effects even during adolescence, as suggested by Sisk & Zehr (2005). Future research might, for example, determine whether the effects of oral contraception on face perception are only acute and reversible, or whether they might be permanent.

6.6.1 The lack of measures of prenatal androgens that would enable further verification of the facial signature

We did not have any direct (prenatal androgens from amniotic fluid) measures of prenatal androgens available in the twin dataset that

would enable us to verify the relationship between prenatal androgens and the facial signature (PC3) there. Future research should consider collaboration with researchers who might have access to measures of prenatal androgens from amniotic fluid as well as MRI data (e.g. Lombardo et al., 2012b) and verify the existence of facial signature.

Subsequently, it is possible that the prediction of relevant disease may be examined in relation to the MRI-derived facial signature of prenatal androgens. Prenatal androgens have been linked with many mental health problems such as autism (Ho et al., 2005), aggression (Cohen-Bendahan et al., 2005), or disordered eating (Culbert et al., 2008). If disease onset could be predicted by both the prenatal androgens (measured from amniotic fluid) and facial signature (skull features of PC3 and related distances) data, the facial signature could be used as an easily accessible indirect index of prenatal androgens and hence the vulnerability to the risk of disease. This could trigger further investigation in this field and potentially lead to earlier diagnosis and disease prevention.

6.6.2 Lack of information that would allow us to predict levels of neurosteroids from serum hormone levels.

Another limitation of our research is the fact that we were studying effects of serum hormone levels on the brain, whereas brain produces its own steroids (neurosteroids) whose levels may be unrelated to hormone levels measured from serum (McCarthy & Konkle, 2005).

Circulating steroid hormones, synthesized in the gonads, adrenal gland, and feto-placental unit, cross the blood-brain barier and can serve as precursors for further synthesis of neurosteroids (Reddy, 2010) that takes place in the hippocampus as well as other brain structures (Baulieu, & Robel, 1990). But astrocytes and neurons are also able to express CYP450scc, which converts cholesterol to pregnenolone and thus initiates the steroid synthesis (Patte-Mensah et al., 2003). Other steroidogenic enzymes in the human brain then might be able to convert pregnenolone to other sex hormones (reviewed in Stoffel-Wanger, 2001). Activity of 5alpha-reductase, which converts testosterone to dihydrotestosterone, was reported as one of the essential rate-limiting steps in synthesis of neurosteroids (Reddy, 2010). Still, the current state of research on human neurosteroids concludes that regulatory mechanisms of neurosteroid synthesis remain unclear (Reddy, 2010).

6.6.3 Sex hormone effects are moderated by genetic factors

Sex differences are determined by the interplay of genes and environment. It is possible that the long-lasting “organizational“ effects of sex steroids might be a result of hormone-induced epigenetic changes in the genome (McCarthy et al., 2009). Studying the epigenetic effects of sex hormones on DNA methylation and histone modification would be an exciting area for future research. This is discussed in more detail below.

As explained in McCarthy et al (2009), epigenetics refers to changes in DNA (DNA methylation) or chromatin (histone modification) that influence gene expression. DNA methylation at gene promoter typically represses gene expression. Histones can either relax or tighten the chromatin surrounding a particular gene and thus modify access to the transcription complex. For example, while estrogen receptor (ER) alpha is highly expressed in the neonatal cortex, this is not the case in the adult neocortex (Previtt, & Wilson, 2007). Westberry et al. (2010) explained that this lack of ER alpha expression in the adult brain is related to DNA methylation. After the neonatal sensitive period, when estradiol could act upon the ER alpha receptor, gene expression of the ER alpha receptor is epigenetically modulated and certain organizational effects on the brain are no longer possible. The fetal programming mechanisms and lasting effects of estradiol on early brain development can thus be explained by epigenetics (McCarthy et al., 2009).

Further research showed that levels of the catalyzers of DNA methylation (DNA cytosine-5-methyltransferases, DNMTs) are sexually dimorphic and responsive to changes in sex hormones (Jessen et al., 2011). This suggests that DNA methylation itself is sexually dimorphic. For example, preoptic area, the brain region important for male sex behavior, has an increased methylation of the ER alpha promoter in adult females, which explains silenced estradiol-responsive sites relevant for male sex behavior (Don Carolos, & Handa, 1994). Kolodkin & Auger (2011) reported that during the first three weeks of life,

amygdala of female rats contains more DNMT3a than male amygdala, which might result in different methylation patterns and program the sex differences in amygdala function and vulnerability to neurodevelopmental disorder such as aggression or anxiety.

According to Lombardo et al. (2012a), individual differences in brain and behavior might be created by epigenetic influences of sex hormones on early brain development (“organizational effects“) that set up the foundation for later interactions of sex hormones with the genome and environment (“activational effects“). Future research on sex differences that aims to apply the mechanisms of the organizational and activational hypothesis should thus consider not only the role of sex hormones but also their interactions with genes and environment.

6.7 CONCLUSIONS

This thesis explored the role of sex hormones as a potential mechanism contributing to sex differences in face development and face processing. Magnetic resonanc images of the head and subsequent MRI-face reconstruction were used to identify craniofacial features that signal one’s sex, age, and exposure to prenatal and postnatal androgens. It was suggested that signature of prenatal androgens in the face might be used, similarly to digit ratio, as an indirect index of prenatal androgens.

The role of sex hormones in face processing was studied with functional MRI and eye-tracking. We demonstrated that use of oral contraception modulates brain response to faces and that duration of oral contraception use predicts plasticity-like changes in the brain. Considering the wide use of oral contraception worldwide, further research about the effects of oral contraception on brain and behavior is imperative.

Overall, we conclude that sex differences in face development and face processing can be, in part, explained by sex hormones. We hope that the further verification of the new indirect index of prenatal androgens will be successful and that facial signature will facilitate more research into the effects of sex hormones and disease risk. Future research should also focus on the interactions between sex hormones, and genes, which will hopefully enable us to explain the presence and implications of sex differences.

REFERENCES

Adamson, P.A., & Zavod, M.B. (2006). Changing perceptions of beauty: A surgeon's perspective. Facial Plastic Surgery, Thieme Medical Publishers Inc., New York. Retrieved Dec 13th 2012 from http://www.dradamson.com/toronto_reconstructive_facial_surgery.html

Ainsworth, M. D. S. (1979). Attachment as related to mother–infant interaction. *Advances in the Study of Behavior, 9,* 2 – 52.

Ankarberg-Lindgren, C., & Norjavaara, E. (2004). Changes of diurnal rhythm and levels of total and free testosterone secretion from pre to late puberty in boys: testis size of 3ml is a transition stage to puberty, *European Journal of Endocrinology, 151,* 747-757.

Antoszewski, B., Zadzinska, E., & Foczpanski, J. (2009). The metric features of teeth in female-to-male transsexuals. *Arch Sex Behav, 38*, 351-358.

Arnold, A.P. (2009). The organizational-activational hypothesis as the foundation for a unified theory of sexual differentiation of all mammalian tissues. *Hormones and Behavior, 55*(5), 570-578.

August, G. P., Grumbach, M. M., & Kaplan, S. L. (1972). Hormonal changes in puberty III: Correlation of plasma testosterone, LH, FSH, testicular size, and bone age with male pubertal development. *Journal of Clinical Endocrinology and Metabolism, 34,* 319-326.

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal*, *12*(1), 26-41.

Baer, H.J., Colditz, G.A., Willett, W.C., & Dorgan, J.F. (2007). Adiposity and sex hormones in girls. *Cancer Epidemiology, Biomarkers, and Prevention, 16,* 1880-1888.

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends Cogn. Sci., 6*, 248–254.

Baron-Cohen, S., Baldwin, D.A., & Crowson, M. (1997). Do children with autism use the speaker's direction of gaze strategy to crack the code of language? *Child Development, 68*(1), 48-57.

Barton, J.J., Press, D.Z., Keenan, J.P., & O'Connor, M. (2002). Lesions of the fusiform face area impair perception of facial configuration in prosopagnosia. *Neurology, 58*(1), 71-78.

Baulieu, E.E, & Robel P. (1990). Neuroactive steroids, a new brain function? *The Journal of Steroid Biochemistry and Molecular Biology, 37*, 395–403.

Bentin, S., Allison, T., Puce, A., Pererz, E., & McCarthy, G. (1996). Electrophysiological studies on face perception in humans. *Journal of Cognitive Neuroscience, 8*, 551-565.

Berenbaum, S.A., Bryk, K.K., Nowak, N., Quigley, C.A., & Moffat, S. (2009). Fingers as a marker of prenatal androgen exposure, *Endocrinology, 150*(11), 5119-5124.

Bindemann, M., Burton, M., Hooge, I.T.C., Jenkins, R., & de Haan, E. (2005). Faces retain attention, *Psychonomic Bulletin and Review, 12*(6), 1048-1053.

Bormann-Kischkel, C., Vilsmeier, M., & Baude, B. (1995). The development of emotional concepts in autism, *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 36*(7), 1243-1259.

Bothwell, R.K., Brigham, J.C.,& Malpass, R.S. (1989). Cross-racial identification. *Personality and Social Psychology Bulletin, 15*, 19-25.

Boyar, R.M., Rosenfeld, R.S., Kapen, S., Finkelstein, J.W., Roffwarg, H.P., Weitzman, E.D., & Hellman, L. (1974). Human puberty: simultaneous augmented secretion of luteinizing hormone and testosterone during sleep. *Journal of Clinical Investigation, 54,* 609-618.

Brown, W.M., Hines, M., Fane, B. A., & Breedlove, S.M. (2002). Masculinized finger length patterns in human males and females with cogenital adrenal hyperplasia, *Hormones and Behavior, 42*, 380-386.

Buchanan, J.R., Hospodar, P., Myers, C., Leuenberger, P., & Demers, L.M. (1988). Effect of excess endogenous androgens on bone density in young women. *J Clin Endocrinol Metab, 67*, 937-943.

Buck, J.J., Williams, R.M., Hughes, I.A., & Acrini, C.L. (2003). In-utero androgen exposure and 2nd to 4th digit length ratio – comparison between healthy controls and females with classical cogenital adrenal hyperplasia, *Human Reproduction, 18*, 976-979.

Bulygina, E., Mitteroecker, P., & Aiello, L. (2006). Ontogeny of facial dimorphism and patterns of individual development within one human population. *Am. J. Phys* Anthropol. 131, 432**–**443.

Carré, J. M., & McCormick, C. M. (2008). In your face: facial metrics predict aggressive behaviour in the laboratory and in varsity and professional hockey players. *Proceedings of the Royal Society B, 275,* 2651-2656.

Chakravarty, M.M, Aleong, R., Leonard,G., Perron,M., Pike, B., Richer, L. et al. (2011). Automated analysis of craniofacialmorphology using magnetic resonance images. *PLoS One 6* (5).

Chapman E, Baron-Cohen S, Auyeung B, Knickmeyer R, Taylor K, & Hackett G. (2006). Fetal testosterone and empathy: evidence from the Empathy Quotient (EQ) and the "reading the mind in the eyes" test. *Soc.Neurosci, 1*, 135–148.

Cheng, Y. D., O'Toole, A. J., & Abdi, H. (2001). Classifying adults' and children's faces by sex: computational investigations of subcategorical feature encoding. *Cognitive Science, 25*, 819-838.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Hillsdale, N.J.: L. Erlbaum Associates.

Cohen-Bendahan, C.C.C., Buitelaar, J.K., van Goozen, S.H.M., Orlebeke, J.F., & Cohen-Kettenis, P.T. (2005). Is there an effect of prenatal testosterone on aggression and other behavioral traits? A study comparing same-sex and opposite-sex twin girls. *Horm. Behav, 47*, 230–237.

Cohen-Bendahan, C.C.C., Buitelaar, J.K., van Goozen, S.M.H., & Cohen-Kettenis, P.T. (2004). Prenatal exposure to testosterone and functional cerebral

lateralization: a study in same-sex and opposite-sex twin girls, *Psychoneuroendocrinology, 29*, 911-916.

Collins, D. L., Holmes, C. J., Peters, T. M., & Evans, A. C. (1995). Automatic 3-D model-based neuroanatomical segmentation. *Human Brain Mapping*, *3*(3), 190-208.

Costa, P.T., Jr., & McCrae, R.R. (1989). *NEO-PI/FFI manual supplement*. Odessa, FL: Psychological Assessment Resources.

Culbert, K.M., Breedlove, M., Burt, A., & Klump, K.L. (2008). Prenatal hormone exposure and risk for eating disorders, *Arch Gen Psychiatry, 65*(3), 329-336.

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

Dagogo-Jack, S., Al-Ali, N., & Qurttom, M (1997). Augmentation of bone mineral density in hirsute women, *J Clin Endocrinol Metab, 82*, 2821-2825.

Dalton, K.M., Nacewicz, B.M., Johnstone, T., Schaefer, H.S., Gernsbacher, M.A., Goldsmith, H.H. et al. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nature Neuroscience, 8*(4), 519-526.

Davidenko, N. (2007). Silhouetted face profiles: a new methodology for face perception research. *Journal of Vision, 7,* 1-17.

Davis EC, Popper P, & Gorski RA. (1996). The role of apoptosis in sexual differentiation of the rat sexually dimorphic nucleus of the preoptic area. *Brain Res, 734*, 10–18.

Dempsey, P.J., Townsend, G.C., & Richards, L.C. (1999). Increased crown size in females with twin brothers: evidence for hormonal diffusion between human twins in utero, *Am J Hum Biol, 11,* 577-586.

Derntl., B., Windischberger, C., Robinson, S., Lamplmayr, E., Kryspin-Exner, I., Gur, R. C. et al. (2008). Facial emotion recognition and amygdala activation are associated with menstrual cycle phase. *Psychoneuroendocrinology, 33*, 1031 – 1040.

Dietrich, T., Krings, T., Neulen, J., Willmes, K., Erberich, S., Thron, A. et al. (2001). Effects of blood estrogen level on cortical activation patterns during cognitive activation as measured by fuctional MRI. *Neuroimage, 13*(3), 425-432.

Diewert, V.M., & Lozanoff, S (1993). A morphometric analysis of human embryonic craniofacial growth in the median plane during primary palate formation. *J Craniofac Genet Dev Biol,13*, 184-192.

Diewert, V.M., Lozanoff, S., & Choy, V. (1993). Computer reconstructions of human embryonic craniofacial morphology showing changes in relations between the face and brain during primary palate formation. *J Craniofac Genet Dev Biol, 13*(3), 193-201.

Domes, G., Sibold, M., Schulze, L., Lischke, A., Herpertz, S.C., & Heinrichs, M. (2012). Intranasal oxytocin increases covert attention to positive social cues, *Psychol Med, 12*, 1-7.

Don Carlos, L.L., & Handa, R.J. (1994). Developmental profile of estrogen receptor mRNA in the preoptic area of male and female neonatal rats. *Brain Res Dev Brain Res, 79*, 283-289.

Dorn, L. D., Nottelmann, E. D., Inoff-Germain, G., Susman, E. J., & Chrousos, G. P. (1990). Perceptions of puberty: adolescent, parent, and health care personnel. *Developmental Psychology, 26*, 322-329.

Dreyer, C.T. (1955). Thoughts on My Craft, Published in *Sight and Sound,* London.

Ejiri, S., Tanaka, M., Watanabe, N., Anwar, R.B., Yamashita, E., Yamada, K. et al. (2008). Estrogen deficiency and its effect on the jaw bones, *J Bone Miner Metab, 26*, 409-415.

Ellenbogen, M.A., Linnen, A.M., Grumet, R., Cardoso, C., & Joober, R. (2012). The acute effects of intranasal oxytocin on automatic and effortful attentional shifting to emotional faces, *Psychophisiology, 49*(1), 128-137.

Engel, A.D., & Haxby, J.V. (2007). Facial expression and gaze-direction in human superior tempoval sulcus, *Neuropsychologia, 45,* 3234-3241.

Enlow, D. (1982). *Handbook of facial growth*. Philadephia, PA: Saunders.

Enlow, D.H., & Hans, M.G.(1996). *Essentials of facial growth*. Saunders Company, Philadelphia, p. 303.

Farah, M.J. (1990). Visual agnosia: Disorders of object recognition and what they tell us about normal vision. Cambridge, MA: MIT Press.

Ferrario, V.F., Sforza, C., Poggio, C.E., & Schmitz, J.H. (1998). Facial volume changes during normal human growth and development. *Anat Record, 250*, 480-487.

Fink, B., Grammer, K., Mitteroecker, P., Gunz, P., Schaefer, K., Bookstein, F.L., & Manning, J.T. (2005). Second to fourth digit ratio and face shape. *Proceedings of the Royal Society of London B: Biological Sciences, 272*, 1995-2000.

Follesa, P., Concas, A., Porcu, P., Sanna, E., Serra, M., Mostallino, M.C. et al. (2001). Role of allopregnanolone in regulation of GABA(A) receptor plasticity during long-term exposure to and withdrawal from progesterone. *Brain Res Rev, 37*(1-3), 81-90.

Fujita T, Kawata T, Tokimasa C, & Tanne K (2001). Influence of oestrogen and androgen on modelling of the mandibular condylar bone in ovariectomized and orchiectomized growing mice. *Arch Oral Biol, 46*, 57-65.

Galsworthy, M.J., Dionne, G., Dále, P.S., & Plomin, R. (2000). Sex differences in early verbal and non-verbal cognitive development, *Dev Sci, 3*, 206-215.

Gauthier, I., Skudlarski, P., Gore, J.C., & Anderson, A.W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition, *Nautre Neuroscience, 3(2).*Geary, D.C. & DeSoto, M.C. (2001). Sex differences in spatial abilities among adults from the United States and China. *Evolution and Cognition, 7*(2). 172-177.

Giedd, J.N. et al. (1999). Brain development during childhood and adolescence; a longitudinal MRI study. *Nature Neuroscience, 2,* 861-863.

Giustina, A., Scalvini, T., Tassi, C., Desenzani, P., Poiesi, C., Wehrenberg, W.B., Rogol, A.D., & Veldhuis, J.D. (1997). Maturation of the regulation of growth hormone secretion in young males with hypogonadotropic hypogonadism pharmacologically exposed to progressive increments in serum testosterone. *Journal of Clinical Endocrinology and Metabolism, 82,* 1210-1219.

Goldberg, S. (2000). *Attachment and development*. London, UK, Hodder Arnold.

Grabner, G., Janke, A. L., Budge, M. M., Smith, D., Pruessner, J., & Collins, D. L. (2006). Symmetric atlasing and model based segmentation: an application to the hippocampus in older adults. *Med Image Comput Comput Assist Interv*, *9*(2), 58-66.

Graham, C.A., Bancroft, J., Doll, H.A., Greco, T., & Tanner, A. (2007). Does oral contraceptive-induced reduction in free testosterone adversely affect the sexuality or mood in women? *Psychoneuroendocrinology, 32*(3), 246-255.

Groppe, S.E., Gossen, A., Rademacher, L., Hahn, A., Westphal, L., Grunder, G., & Spreckelmeyer, K.N. (2013). Oxytocin influences processing of socially relevant cues in the ventral tegmental area of the human brain, *Biological Psychiatry.*

Grosbras, M.H., & Paus, T. (2006). Brain networks involved in viewing angry hands or faces. *Cerebral Cortex, 16*(8), 1087-1096.

Guihard-Costa, A.M., Khung, S., Delbecque, K., Menez, F., & Delezoide, A.L. (2006). Biometry of face and brain in fetuses with trisomy 21. *Pediatric Research, 59*, 33-38.

Hall J.A., & Matsumoto, D. (2004). Gender differences in judgments of multiple emotions from facial expressions, *Emotion, 4*(2): 201-206.

Hall, J. A., Lamb, E., & Perlmutter, M. (1986). *Child psychology today (2nd ed.).* New York, Random House.

Hall, J.A., & Matsumoto, D. (2004). Gender differences in judgments of multiple emotions from facial expressions. *Emotion, 4*(2), 201–6.

Hall, J.K., Hutton, S.B., & Morgan, M.J. (2010). Sex differences in scanning faces: Does attention to the eyes explain female superiority in facial expression recognition? *Cognition and Emotion, 24*(4), 629-637.

Halpern, D. F. & Tan, U. (2001). Stereotypes and steroids: using a psychobiosocial model to understand cognitive sex differences, *Brain Cogn,* 45, 392–414.

Hampson, E., & Kimura, D. (1988). Reciprocal effects of hormonal fluctuations on human motor and perceptual-spatial skills. *Behavioral Neuroscience, 102*(3), 456-459.

Hampson, E., & Young, E.A. (2007). Chapter 4: Methodological Issues in the Study of Hormone-Behavior Relations in Humans: Understanding and Monitoring the Menstrual Cycle. In Becker, J.B., Berkley, K.J., Geary, N., Hampson, E., Herman, J.P., & Young, E. (2007). *Sex Differences in the Brain: From Genes to behavior.* New York: Oxford University Press.

Hampson, E., Vanandres, S., & Mullin, L. (2006). A female advantage in the recognition of emotional facial expressions: test of an evolutionary hypothesis. *Evolution and Human Behavior, 27*(6): 401-416.

Haxby, J.V., Hoffman, E.A., & Gobbini, M.I. (2002). The distributed human neural system for face perception. *Trends in Cognitive Science, 4*, 223-333.

Herba, C., & Phillips, M. (2004). Annotation: Development of facial expression recognition from childhood to adolescence: behavioural and neurological perspectives, *Journal of Child Psychology and Psychiatry, 45*(7), 1185-1198.

Hariri, A. R., Mattay, V.S., Tessitore, A., Kolachana, B., Fera, F., Goldman, D. et al. (2002). Serotonin transporter genetic variation and the response of the human amygdala, *Science, 400.*

Herlitz, A., & Rehnman, J. (2008). Sex differences in episodic memory. *Current Directions in Psychological Science, 17(*1), 52-56.

Herlitz, A., Nilsson, L.G., & Backman, L. (1997). Gender differences in episodic memory. *Memory & Cognition, 25*, 801-811.

Hietala, M., Sandberg, T., Borg, A., Olsson, H., & Jenstrom, H. (2007). Testosterone levels in relation to oral contraceptive use and the androgen receptor CAG and GGC length polymorphisms in healthy young women. *Human Reproduction, 22*(1), 83-91.

Hines, M. (1990). Gonadal hormones and human cognitive development. In J. Balthazart (Ed.), *Brain and behavior in vertebrates 1: Sexual differentiation, neuroanatomical aspects, neurotransmitters, and neuropeptides*, Basel, Schwitzerland: Karger, 51-63.

Ho, A., Todd, R.D., & Constantino, J.N. (2005). Brief report: autistic traits in twins vs. nontwins a preliminary study. *J. Autism Dev. Disord. 35*, 129–133.

Hobson, R.P., & Lee, A. (1998). Hello and goodbye: a study of social engagement in autism. *J Autism Dev Disord, 28*, 117-127.

Hoffmann, H., Kessler, H., Eppel, T., Rukavina, S., & Traue, H.C. (2010). Expression intensity, gender and facial emotion recognition: Women recognize only subtle facial emotions better than men. *Acta Psychologica, 135*(3), 278-283.

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

Hoss, R.A. et al (2005). The role of facial attractiveness and facial masculinity/femininity in sex classification of faces, *Perception, 34*, 1459-1474.

Hyde, J.S., & Linn, M.C. (1988). Gender differences in verbal ability: a meta-analysis, *Psychological Bulletin, 104*(1), 53-69.

Irvine, R.A., Ma, H., Yu, M.C., Ross, R.K., Stallcup, M.R., & Coetzee, G.A. (2000). Inhibition of p160-mediated coactivation with increasing androgen receptor polyglutamine length. *Human Molecular Genetics, 9,* 267-274.

Jessen HM, & Auger AP (2011): Sex differences in epigenetic mechanisms may underlie risk and resilience for mental health disorders. *Epigenetics, 6*, 857–861.

Kanwisher, N., McDermott, J., & Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience, 17*(11), 4302-4311.

Kanwisher, N., Stanley, D., & Harris, A. (1999). The fusiform face area is selective for faces not animals. *Neuroreport, 10*, 183-187.

Kashyap, R. R., Babu, G. S., & Shetty, S. R. (2011). Dental patient with acromegaly: a case report. *A Journal of Oral Science, 53*, 133-136.

Kastner, S., Pinsk, M.A., De Weerd, P., Desimone, R., & Ungerleider, L.G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron, 22*, 751-761.

Keenan, B.S., Richards, G.E., Ponder, S.W., Dallas, J.S., Nagamani, M., & Smith, E.R. (1993). Androgen-stimulated pubertal growth: the effects of testosterone and dihydrotestosterone on growth hormone and insulin-like growth factor-I in the treatment of short stature and delayed puberty. *Journal of Clinical Endocrinology and Metabolism, 76,* 996-1001.

Kimura, D. (1999). *Sex and Cognition*. Cambridge: MIT Press.

Kirouac, G. & Doré, F.Y. (1985). Accuracy of the judgment of facial expression of emotions as a function of sex and level of education. *Journal of Nonverbal Behavior, 9*, 3-7.

Knickmeyer, R., Baron-Cohen, S., Raggatt, P., Taylor, K., & Hackett, G. (2006). Fetal testosterone and empathy. *Hormones and Behaviour, 49*(3), 282_292.

Knickmeyer, R.C., Wheelwright, S., Taylor, K., Raggatt, P., Hackett, G., & Baron-Cohen, S. (2005). Gender-typed play and amniotic testosterone. *Dev. Psychol. 41*, 517–528.

Kolodkin, M.H., & Auger, A.P. (2011). Sex differences in the expression of DNA methyltransferase 3a (DNMT3a) in the rat amygdala during development. *J Neuroendocrinol,* In press.

Kramer, K.M., Cushing, B.S., Carter, C.S., Wu, J., & Ottinger, M.A. (2004). Sex and species differences in plasma oxytocin using an enzyme immunoassay, *Can J Zool, 82*, 1194-1200.

Lane, R.D., Chua, P.M., & Dolan, R.J. (1999). Common effects of emotional valence, arousal and attention on neural activation during visual processing of pictures, *Neuropsychologia, 37*(9), 989-997.

Larkby, C., & Day, N. (1997). The effects of prenatal alcohol exposure. *Alcohol health & research world, 21*(3), 192-198.

Lawrence-Owen, A.J., Bargar, G., Bosten, J.M., Goodbourn, P.T., Hogg, R.E., & Mollon, J.D. (2012). Genetic association suggests that SMOC1 mediates between prenatal sex hormones and digit ratio. *Human genetics.*

Lazic, M., Aird, F., Levine, J.E., & Dunaif, A. (2011). Prenatal androgen treatment alters body composition and glucose homeostasis in male rats, *Journal of Endocrinology, 208*, 293-300.

Lewin, C., Wolgers, G. & Herlitz, A. (2001). Sex differences in face recognition – Women's faces make the difference. *Brain and Cognition, 50*(1), 121-128.

Lilley, T., Laaksonen, T., Huitu, O., & Helle, S. (2010). Maternal corticosterone but not testosterone level is associated with the ratio of second-to-fourth digit length (2D:4D) in field vole offspring (Microtus agrestis). *Physiol Behav, 99*(4).

Lischke, A., Berger, C., Prehn, K., Heinrichs, M., Herpertz, S.C., & Domes, G. (2012). Intranasal oxytocin enhances emotion recognition from dynamic facial expressions and leaves eye-gaze unaffected, *Psychoneuroendocrinology, 37*(4), 475-481.

Liu F, van der Lijn F, Schurmann C, Zhu G, Chakravarty MM, et al. (2012) A Genome-Wide Association Study Identifies Five Loci Influencing Facial Morphology in Europeans. *PLoS Genet 8*(9), 1-13.

Loche, S., Colao, A., Cappa, M., Bellone, J., Aimaretti, G., Farello, G., Faedda, A., Lombardi, G., Deghengui, R., & Ghigo, E. (1997). The growth hormone response to hexarelin in children: reproducibility and effect of sex steroids. *Journal of Clinical Endocrinology and Metabolism, 82,* 861-864.

Lombardo, M.V., Ashwin, E., Auyeung, B., Chakrabarti, B., Lai, M.C., Taylor, K., Hackett, G., Bullmore, E.T., & Baron-Cohen, S. (2012a). Fetal programming effects of testosterone on the reward system and behavioral approach tendencies in humans, *Biological Psychiatry, 72,* 839-847.

Lombardo, M.V., Ashwin, E., Auyeung, G., Chakrabarti, B., Taylor, K., Hackett, G., Bullmore, e.t., & Baron-Cohen, S. (2012b). Fetal testosterone influences sexually dimorphic gray matter in the human brain, *Journal of Neuroscience, 11*, 674-680.

Lutchmaya, S., Baron-Cohen, S., Raggatt, P., Knickmeyer, R., & Manning, J.T. (2004). 2nd to 4th digit ratios, fetal testosterone and estradiol, *Early Human Development, 77*, 23-28.

Macrae, C.N., Alnwick, K.M., Milne, A.B., & Scholerscheidt, A.M. (2002). Person perceptions across menstrual cycle. Hormonal influences on social-cognitive funcitoning. *Psychological Science, 13*, 532-536.

Maki, P.M., Rich, J.B., & Rosenbaum, R.S. (2002). Implicit memory varies across the menstrual cycle: estrogen effects in young women, *Neuropsychologia, 40*(5), 518-529.

Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology, 13*, 330-334.

Manning, J. T., Scutt, D., Wilson, J. & Lewis-Jones, D.I. (1998). The ratio of 2^{nd} to 4^{th} digit length: a predictor of sperm numbers and concentrations of testosterone, luteinizing hormone and oestrogen, *Human Reproduction, 13*, 3000-3004.

Manning, J.T., Churchill, J.G.A., & Peters, M. (2007). The Effects of Sex, Ethnicity, and Sexual Orientation of Sef-Measured Digit Ratio (2D:4D). *Archives of Sexual Behavior, 36*(2), 223-233.

Manning, J.T., Stewart, A., Bundred, P.E., & Trivers, R.L. (2004). Sex and ethnic differences in 2^{nd} to 4^{th} digit ratio of children. *Early Human Development, 80(*2), 161-168.

Mareckova, K., Weinbrand, Z., Chakravarty, M.M., Lawrence, C., Aleong, R. et al. (2011). Testosterone-mediated sex differences in the face shape during adolescence: Subjective impressions and objective features. *Hormones and Behavior, 60,* 681-690.

Martin, D., & Macrae, C. N. (2007). A face with a cue: exploring the inevitability of person categorization. *European Journal of Social Psychology, 37,* 806-816.

Mathews, G.A., Fane, B.A.,Conway, G.S., Brook, C., & Hines, M. (2009). Personality and congenital adrenal hyperplasia: possible effects of prenatal androgen exposure. *Horm. Behav. 55*, 285–91

McBain, R., Norton, D., & Chen, Y. (2009). Females excel at basic face perception. *Acta Psychol (Amst), 130*(2): 168-73.

McCarthy MM, Auger AP, Bale TL, De Vries GJ, Dunn GA, Forger NG, et al. (2009): The epigenetics of sex differences in the brain. *J Neurosci, 29*, 12815–12823.

McCarthy, M.M. (1995). Estrogen modulation of oxytocin and its relation to behavior. *Adv. Exp Med Biol., 395*, 235-245.

McCarthy, M.M., & Konkle, A.T.M. (2005). When is a sex difference not a sex difference? *Frontiers in Neuroendocrinology, 26(*2), 85-102.

McCartney, C. R., Blank, S. K., Prendergast, K. A., Chhabra, S., Eagleson, C. A., Helm, K. D. et al. (2007). Obesity and sex steroid changes across puberty: evidence for marked hyperandrogenemia in pre- and early pubertal obese girls. *J Clin Endocrinol Metab, 92*(2), 430-436.

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.

McIntyre, M.H. (2006). The use of digit ratios as markers for perinatal androgen action. *Reproductive Biology and Endocrinology, 4*(10).

McGugin, R.W., Gatenby, J.C., Gore, J.C., & Gauthier, I. (2012). High-resolution imaging of expertise reveals reliable object selectivity in the fusiform face area related to perceptual performance. PNAS, 109(42).

Meaney, M. J., & Stewart, J. (1981). Neonatal androgens influence the social play of prepubescent male and female rats. Hormones & Behavior, 15, 197-213.

Medical dictionary (2011). Retrieved November 20, 2012, from http://de.dict.md/definition/estradiol

Medland, S.E., Zayats, T., Glaser, B., Nyholt, D.R., Gordon, S.D., Wright, M.J. et al. (2010). A variant in LIN28B is associated with 2D4D finger-length ratio, a putative retrospective biomarker of prenatal testosterone exposure. *Am J Hum Gen, 86*(4), 519-525.

Meindl, K., Windhager, S., Wallener, B., & Schaefer, K. (2012). Second-to-fourth digit ratio and facial shape in boys: the lower the digit ratio, the more robust the face. *Proc R Soc B, 279*(1737), 2457-2463.

Miyahira, A., Morita, K., Yamaguchi, H., Nonaka, K., & Maeda, H. (2000). Gender differences of exploratory eye movements: a life span study. *Life Sci, 68*(5), 569-77.

Montagne, B., Kessels, R.P.C., Frigerio, E., de Haan, E.H.F., & Perrett, D.I. (2005). Sex differences in the perception of affective facial expressions: do men really lack emotional sensitivity? *Cogn Process, 6*(2): 136-141.

Mordecai, K.L., Rubin, L.H., & Maki, P.M. (2008). Effects of menstrual cycle phase and oral contraceptive use on verbal memory. *Hormones and Behavior, 54*(2), 286-293.

Morishima A, Grumbach MM, Simpson ER, Fisher C, & Qin K (1995). Aromatase deficiency in male and female siblings caused by a novel mutation and the physiological role of estrogens*. J Clin Endocrinol Metab, 80*, 3689-3698.

Morris, J.P., Pelphrey, K.A., & McCarthy, G. (2007). Controlled scanpath variation alters fusiform face activation. *Social Cognitive and Affective Neuroscience, 2*(1), 31-38.

Mosher, W.D., Martinez, G.M., Chandra, A., Abma, J.C., & Willson, S.J. (2004). Use of contraception and use of family planning services in the United States: 1982 – 2002. *Adv Data, 350*, 1-36.

Mueller U., & Mazur A (1996). Facial dominance of West Point cadets as a predictor of later military rank, *Soc. Forces, 74***,** 823–850.

Murakami S, Arai Y. (1989). Neuronal death in the developing sexually dimorphic periventricular nucleus of the preopti area in the female rat: effect of neonatal androgen treatment. *Neurosci Lett, 102*, 185–190.

Naftolin, F., Ryan, K.J., Davies, I.J., Reddy, W., Flores, F., Petro, Z. et al. (1975). The formativ of estrogens by central neuroendocrine tissues. *Recent Prog Horm Res, 31*, 295-319.

Neave, N., Laing, S., Fink, B., & Manning, J.T. (2003). Second to fourth digit ratio, testosterone and perceived male dominance, *Proc R Soc Lond B, 270*, 2167-2172.

Nunez, J.L., Sodhi, J., & Juraska, J.M. (2002). Ovarian hormones after postnatal day 20 reduce neuron number in the rat primary visual cortex, *J. Neurobiol., 52*, 312–321.

Okten, A., Kalyoncu, M., & Yaris, N. (2002). The ratio of second- and fourth-digit lengths and congenital adrenal hyperplasia due to 21-hydroxylase deficiency, *Early Hum Dev, 70*(1-2), 47-54.

O'Toole, A. J., Deffenbacher, K. A., Valentin, D., McKee, K., & Abdi, H. (1998). The perception of face gender: the role of stimulus structure in recognition and classification. *Mem Cognit, 26* (1), 146-160.

Palermo, R., & Rhodes, G. (2007). Are you always on my mind? A review of how face perception and attention interact. *Neuropsychologia, 45*(1), 75-92.

Parducz, A., Hoyk, Z., Kis, Z., & Garcia-Segura, L.M. (2002). Hormonal enhancement of neuronal firing is linked to structural remodelling of excitatory and inhibitory synapses. *European Journal of Neuroscience, 16*(4), 665-670.

Pasterski, V., Hindmarsh, P., Geffner, M., Brook, C., Brain, C., & Hines, M. (2007). Increased aggression and activity level in 3- to 11-year-old girls with congenital adrenal hyperplasia (CAH), *Hormones and Behavior, 52*(3), 368-374.

Patte-Mensah, C., Kappes, V., Freund-Mercier, M.J., Tsutsui, K., & Mensah-Nyagan, AG. (2003). Cellular distribution and bioactivity of the key steroidogenic enzyme, cytochrome P450side chain cleavage, in sensory neural pathways. *Journal of Neurochemistry, 86*, 1233–1246.

Paus, T. (2013). *Population Neuroscience.* Springer-Verlag Berlin Heidelberg 2013, ISBN-10: 3642364497.

Pausova, Z., Paus, T., Abrahamowicz, M., Almerigi, J., Arbour, N., Bernard, M. et al. (2007). Genes, maternal smoking, and the offspring brain and body during adolescence: Design of the Saquenay Youth Study. *Human Brain Mapping, 28,* 502-518.

Pearson, T., & Lewis, M.B. (2005). Fear recognition across the menstrual cycle. *Hormones and Behavior, 47*, 267–271.

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

Penton-Voak, I.S., & Perrett, D.I. (2000). Female preference for male faces changes cyclically: Further evidence. *Evolution and Human Behavior, 21*, 39-48.

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

Penton-Voak, I.S., Jones, B.C., Little, A.C., Baker, S., Tiddeman, B., Burt, D.M., & Perret, D.I. (2001). Symmetry, sexual dimorphism in facial proportions and male facial attractiveness. *Proc. Royal Soc. London, 268,* 1617–1623.

Peper, J.S., Brouwer, R.M., van Baal, G., Schnack, H., leeuwen, M., Boomsma, D., Kahn, R., & Pol, H. (2009). Does having a twin brother make for a bigger brain? *Eur. J. Endocrinol, 160*, 739-746.

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

Perrin, J.S,, Leonard, G., Perron, M., Pike, G.B., Pitiot, A., Richer, L., Veillette, S., Pausova, Z., & Paus, T. (2009). Sex differences in the growth of white matter during adolescence, *Neuroimage, 45*(4), 1055-1066.

Perrin, J.S., Hervé, P.-Y., Leonard, G., Perron, M., Pike, G.B., Pitiot, A., Richer, L., Veillette, S., Pausova, Z., & Paus, T. (2008). Growth of white matter in the adolescent brain: role of testosterone and androgen receptor. *Journal of Neuroscience, 28,* 9519-9524.

Peterson, 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.

Petrovic, P., Kalisch, R., Singer, T., & Dolan, R. J. (2008). Oxytocin attenuates affective evaluations of conditioned faces and amygdala activity. *J Neurosci, 28*(26), 6607-6615.

Phoenix, C.H., Goy, R.W., Gerall, A.A., & Young, W.C. (1959). Organizing action of prenatally administered testosterone propionáte on the tissue mediating mating behavior in the female guinea pig. *Endocrinology, 65*, 369-382.

Pierce, K., & Schreibman, L. (1995). Increasing komplex social behaviors in children with autism: effects of peer implemented pivotal response training. *J Appl Behav Anal, 28*, 285-295.

Pierce, K., Muller, R.A., Ambrose, J., Allen, G., & Courchesne, E. (2001). Face processing occurs outsider the fusiform face area in autism: evidence from functional MRI. *Brain, 124*, 2059-2073.

Pierce, M. (2012). Facial Expression Intelligence Scale (FEIS): Recognizing and Interpreting Facial Expressions and Implications for Consumer Behavior. Virginia Polytechnic Institute and State University, Virginia.

Pinos, H., Collado, P., Rodriguez-Zafra, M., Rodriguez, C., Segovia, S., & Guillamon, A. (2001). The development of sex differences in the locus coeruleus of the rat, *Brain Res. Bull, 56,* 73–78.

Posamentier, M.T., & Abdi, H. (2003). Processing faces and facial expressions. *Neuropsychology Review, 13*(3), 113-143.

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

Preacher, K.J., & Hayes, A.F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers, 36* (4), 717-731.

Prewitt A.K., & Wilson M.E. (2007). Changes in estrogen receptor- mRNA in the mouse cortex during development. *Brain Res, 1134*, 62–69.

Rahman, Q., Wilson, G.D., & Abrahams, S. (2004). Sex, sexual orientation, and identification of positive and negative facial affect, *Brain and Cognition, 54*(3), 179-185.

Rainey, W. E., Rehman, K. S., & Carr, B. R. (2004) The human fetal adrenal: making adrenal androgens for placental estrogens. *Semin Reprod Med,* 22, 327-336.

Reddy, D.S. (2010). Neurosteroids: Endogenous role in the human brain and therapeutic potentials, *Prog Brain Res, 186*, 113-137.

Rehnman, J., & Herliz, A. (2006). Higher face recognition ability in girls: magnified by own-sex and own-ethnicity bias. *Memory, 14,* 289-296.

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology, 57*, 199–226.

Riggs, B.L., Khosla, S., & Melton, L.J. (2002). Sex steroids and the construction and conservation of the adult skeleton, *Endocr Rev, 23*, 279-302.

Rosenberg, L., & Park, S. (2002). Verbal and spatial functions across the menstrual cycle in healthy young women. *Psychoneuroendocrinology, 27*, 835-841.

Rosenfield, R. L. (2007). Clinical review: Identifying children at risk for polycystic ovary syndrome. *J Clin Endocrinol Metab, 92*(3), 787-796.

Sadeghianrizi, A., Forsberg, C.-M., Marcus, C., & Dahllöf, G. (2005). Craniofacial development in obese adolescents. *European Journal of Orthodontics, 27*, 550-555.

Sakai, L.M., Baker, L.A., Jacklin, C.N., & Shulman, I. (1992). Sex steroids at birth: genetic and environmental variation and covariation, *Developmental Psychobiology*, 24, 559-570.

Sandres, G., Sjodin, M., & de Chastelaine, M. (2002). On the elusive nature of sex differences in cognition: Hormonal influences contributing to within-sex variation. *Archives of Sexual Behavior, 31*, 145-152.

Sarkar, P., Bergman, K., Fisk, N. M., O'Connor, T. G., & Glover, V. (2007) Amniotic fluid testosterone: relationship with cortisol and gestational age. *Clin Endocrinol (Oxf),* 67, 743-747.

Schaefer, K., Fink, B., Mitteroecker, P., Neave, N., & Bookstein, F.L. (2005). Visualizing facial shape regression upon 2nd to 4th digit ratio and testosterone. *Collegium Anthropologicum, 29*(2), 415-419.

Scholten, M.R.M., Aleman, A., Montagne, B., & Kahn, R.S. (2005). Schizophrenia and processing of facial emotions: sex matters. *Schizophrenia Research, 78*(1), 61-67.

Schulte-Ruther, M., Markowitch, H.J., Shah, J.N., Fink, G.R., & Piefke, M. (2008). Gender differences in brain networks supporting empathy. *NeuroImage, 42,* 393-403.

Schultz, R.T. (2005). Developmental deficits in social perception in autism: the role of the amygdala and fusiform face area, *Int. J. Devl Neuroscience, 23*, 125–141.

Schulz, K.M., Molenda-Figueira, H.A., & Sisk, C.L. (2009). Back to the future: The organizational-activational hypothesis adapted to puberty and adolescence. *Hormones and Behavior, 55*(5), 597-604.

Schumann G, Loth E, Banaschewski T, Barbot A, Barker G, Buchel C et al. and the IMAGEN consortium (2010). The IMAGEN study: Reinforcement-related behaviour in normal brain function and psychopathology. *Molecular Psychiatry, 15*(12), 1128-1139.

Scott I.M.L. et al. (2010) Does Masculinity Matter? The contribution of masculine face shape to male attractiveness in humans. *PLoS ONE 5*(10).

Silber, M., Almkvist, O., Larsson, B., Stock, S., & Uvnas-Moberg, K. (1987). The effect of oral contraceptive pills on levels of oxytocin in plasma and on cognitive functions. *Contraception, 36* (6), 641-50.

Sisk, C.L., & Zehr, J.L. (2005). Pubertal hormones organize the adolescent brain and behavior. *Front Neuroendocrinol.,* 26,163–174.

Sled, J. G., Zijdenbos, A. P., & Evans, A. C. (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imaging*, *17*(1), 87-97.

Smith, S.S. (1994). Female sex steroid hormones: from receptors to networks to performance – actions on the sensorimotor system. *Progress in Neurobiology, 44*, 55-86.

Society for Biomedical Diabetes Research (2010). SI Unit Conversion Calculator. Retrieved November 20, 2012 from http://www.socbdr.org/rds/authors/unit_tables_conversions_and_genetic_dictionaries/e5196/index_en.html

Södergård, R., Bäckström, T., Shanbhag, V., & Cartensen, H. (1982). Calculation of free and bound fractions of testosterone and estradiol-17 beta to human plasma proteins and body temperatures. *Journal of Steroid Biochemistry, 16,* 801-810.

Sperber, G.H., Gutterman, G.D., & Sperber, S.M. (2000). *Craniofacial development.* Hamilton, Ont., B.C. Decker Inc

Spiegel, R., Sather, H., & Hayles, A. (1971). Cephalometric study of children with various endocrine diseases. *American Journal of Orthodontics, 59*, 362-375.

Stirrat, M.,& Perret, D.I. (2010). Valid facial cues to cooperation and trust: male facial width and trusworthiness, *Psychol Sci., 21* (3), 349-54.

Stock, S., Karlsson, R., & von Schoultz, B. (1994). Serum profiles of oxytocin during oral contraceptive treatment. *Gynecol Endocrinol, 8*(2), 121-126.

Stoffel-Wagner, B. (2001). Neurosteroid metabolism in the human brain, European Journal of Endocrinology, 145, 669-679.

Sullivan, S., Ruffman, T., & Hutton, S.B. (2007). Age differences in emotion recognition skills and the visual scanning of emotion faces, *Journals of Gerontology: Series B, 62*(1), 53-60.

Tahmasebi, A., Artiges, E., Banaschewski, T., et al. (2012). Creating probabilistic maps of the face network in the adolescent brain: a multi-centre functional MRI study. *Human Brain Mapping, 33*(4), 938–57.

Tanaka, M., Ejiri, S., Nakajima, M., Kohno, S., & Ozawa, H. (1999). Changes of cancellous bone mass in rat mandibular condyle following ovariectomy, *Bone, 25*, 339-347.

Tapp, A. L., Maybery, M.T., & Whitehouse, A.J.O. (2011). Evaluating the twin testosterone transfer hypothesis: A review of the empirical evidence. *Hormones and Behavior, 60*, 713-722.

Thordarson, A., Johannsdottir, B., & Magnusson, T.E. (2006). Craniofacial changes in Icelandic children between 6 and 16 years of age – a longitudinal study. *European Journal of Orthodontics, 28*, 152-165.

Thornhill, R., & Gangestad, S.W. (1999). Facial attractiveness. *Trends Cogn Sci, 3*, 452–460.

Todorov & Engell (2008).The role of amygdala in implicite evaluation of emotionally neutral faces, *Social Cognitive and Affective Neuroscience, 3*, 303-312.

Toma A.M. et al (2008). A three-dimensional look for facial differences between males and females in a British-Caucasian sample aged 15½ years old. *Orthodontics and Craniofacial Research 11*(3):180-185.

Toran-Allerand, C.D. (1976). Sex steroids and the development of the newborn mouse hypothalamus and preoptic area in vitro: implications for sexual differentiation. *Brain Res, 106*, 407–412.

Van Anders & Verhon, P.A. (2006). Finger-length ratios show evidence of prenatal hormone transfer between opposite-sex twins, *Hormones and Behavior, 49*, 315-319.

Van de Beek, C., Thijssen, J.H.H., Cohen-Kettenis, P.T., van Goozen, S.H.M., & Buitelaar, J.K. (2004). Relationships between sex hormones assessed in amniotic fluid, and maternal and umbilical cord serum: what is the best source of information to investigate the effects of fetal hormonal exposure? *Horm. Behav., 46*, 663–669.

Van den Heuvel, M.W., van Bragt, A.J., Alnabawy, A.K., & Kaptein, M.C. (2005). Comparison of ethinylestradiol pharmacokinetics in three hormonal

contraceptive formulations: the vaginal ring, the transdermal patch and an oral contraceptive. *Contraception, 72*(3), 168-174.

Van Honk, J., Schutter, D. J., Bos, P. A., Kruijt, A.W., Lentjes, E. G., & Baron-Cohen, S. (2011). Testosterone administration impairs cognitive empathy in women depending on second-to-fourth digit ratio. *Proceedings of the National Academy of Sciences, 108*, 3448–3452.

Vanderschueren, D., Vandenput, L., Boonen, S., Lindberg, M.K., Bouillon, R., & Ohlsson, C. (2004). Androgens and bone, *Endocrine Reviews, 25*(3), 389-425.

Verdonck, A., Gaethofs, M., Carels, C., & de Zegher, F. (1999). Effect of low-dose testosterone treatment on craniofacial growth in boys with delayed puberty. *European Journal of Orthodontics, 21*, 137-143.

Voracek, M., & Dressler, S. G. (2006). Lack of correlation between digit ratio (2D:4D) and Baron-Cohen's "reading the mind in the eyes" test, empathy, systemising, and autism-spectrum quotients in a general population sample. *Personality and Individual Differences, 41*, 1481–1491.

Voyer, D., Voyer, S. & Bryden, M.P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117*(2), 250-270.

Vuoksimaa, E., Kaprio, J., Kremen, W.S., Hokkanen, L., Viken, R.J., & Tuulio-Henriksson, A. (2010). Having a male co-twin masculinizes mental station performance in females. *Psychological Science, 21*(8), 1069-1071.

Wallen K., & Baum, MJ. (2002). Masculinization and defeminization in altricial and precocial mammals: comparative aspects of steroid hormone action. In: Pfaff DW, Arnold AP, Etgen AM, Fahrbach SE, Rubin RT, editors. *Hormones, Brain and Behavior.* San Diego: Elsevier Science (USA), 385–423.

Wassermann E, Epstein, C.M., Ziemann U, Walsh, V., Paus, T., & Lisanby, S.H. (Eds). (2008) *The Oxford Handbook of Transcranial Stimulation*. Oxford, Oxford University Press.

Westberry JM, Trout AL, & Wilson ME (2010). Epigenetic regulation of estrogen receptor- gene expression in the mouse cortex during early postnatal development. *Endocrinology, 2*, 731–740.

Weston, E.M., Friday, A.E., Johnstone, R.A., & Schrenk, F. (2004). Wide faces or large canines? The attractive versus aggressive primate. *Proc R Soc London Ser B Suppl Biology Letters*, 271, S416-S419.

Weston, E.M., Friday, A.E., & Liò, P. (2007). Biometric evidence that sexual selection has shaped the Hominin face. *PLoS One 2*, 1**–**8.

Wharton, W., Hirshman, E., Merritt, P., Doyle, L., Paris, S., & Gleason, C. (2008). Oral contraceptives and androgenicity: influences on visuospatial task performace in younger individuals. *Exp Clin Psychopharmacology, 16*(2), 156-64.

Wild, H., Barett, S., Spence, M., O'Toole A., Chenh, Y., & Brooke, J. (2000). Recognition and sex classification of adults' and children's faces: examining performance in the absence of sex-stereotyped cues. *Journal of Experimental Child Psychology, 77*, 269-291.

Zijdenbos, A. P., Forghani, R., & Evans, A. C. (2002). Automatic "pipeline" analysis of 3-D MRI data for clinical trials: application to multiple sclerosis. *IEEE Trans Med Imaging*, *21*(10), 1280-1291.

Zingeser, M.R., & Phoenix, C.H. (1978). Metric characteristics of the canine dental komplex in prenatally androgenized female rhesus monekys (Macaca mulatta), *Am J Phys Anthrop, 49*, 187-192.

Zingg, H.H. (2002). Oxytocin. In: Pfaff, D. (Ed) et al. *Hormones, brain and behavior, vol. III.* Academic Press, San Diego, 779-802.

Zuolaga, D. (2008). *The role of the androgen receptor in anxiety-related behaviors, the hypothalamic pituitary adrenal axis, and sensorimotor gating: studies in rodents with the testicular feminization mutation*, Michigan State University.

APPENDICES

Appendix 1: Marečková, K., Weinbrand, Z., Chakravarty, M.M., Lawrence, C., Aleong, R., Leonard, G., Perron, M., Pike, G.B., Richer, L., Veillette, S., Pausova, Z., & Paus, T. (2011). Testosterone-mediated sex differences in the face shape during adolescence: Subjective impressions and objective features. *Hormones and Behavior, 60,* 681-690.

Appendix 2: Marečková, K., Chakravarty, M.M., Huang, M., Lawrence, C., Leonard, G., Perron, M., Pike, B.G., Richer, L., Veilette, S., Pausova, Z., & Paus, T. (2013). Does skull shape mediate the relationship between objective features and subjective impressions about the face? *Neuroimage, 79*, 234-240.

Appendix 3: Marečková, K., Perrin, J.S., Nawaz-Khan, I., Lawrence, C., Dickie, E., McQuiggan, D.A., Paus, T., & the IMAGEN Consortium (2012). Hormonal contraceptives, menstrual cycle and brain response to faces, *Social Cognitive and Affective Neuroscience*. PMID: 23175677

Appendix 4: Effect of rater's sex on accuracy of face sex recognition

Appendix 5: Distances between face landmarks

Appendix 6: Distances between skull landmarks

Appendix 7: Relationship between the face PCs described in Chapter 3 (current) and face PCs described in Chapter 2 (Mareckova et al., 2011).

Appendix 8: Effect of menstrual cycle and oral contraception use on amygdala BOLD response to faces.

APPENDIX 4

Effect of rater's sex on accuracy of face sex recognition

To assess raters' accuracy in identifying the sex of face stimuli, we used d' ('d-prime'), a robust measure widely used in signal-detection literature to determine the sensitivity by which raters make decisions about stimuli. This is considered to be a pure measure of raters' ability to distinguish between signal (in this case, the actual sex of the face presented in a given trial) and noise (the alternative possible sex of a given presented face). D-prime is calculated as the difference between the z-score of the hit rate and the z-score of the false-alarm rate. Scores of d' range from 0 to 1, where d' of 0 indicates the poorest ability to distinguish between signal and noise, and 1 indicates the strongest ability to distinguish between signal and noise.

The d' scores of female raters (M = 0.59, SD = 0.18) were significantly higher than those of male raters (M = 0.41, SD = 0.21; t (35) = 2.95, p = 0.005). The effect size (d) of this sex difference was 0.95, and is considered large by Cohen (Cohen, 1988). There was no significant interaction between the sex of the rater and the objective sex of the face (p = 0.74) and thus the possibility that responses from male and female raters would show an opposite direction is ruled out. Better female performance in identifying sex of faces is consistent with previous reports (e.g., Rehnman & Herlitz, 2006). Given the superiority of the female raters and our interest in the most accurate judgments, we carried out all subsequent analyses using ratings made only by females. Thus, the group of

raters we will refer to from now on were 60 White Caucasian females whose mean age was 18.7 years, SD = 1.21.

Sex rating of the average face

The average face was created using an approximately equal proportion of male and female faces of equal age distributions.

Each presentation of the model face was considered as an independent event. This allowed us to evaluate this average face image as if it were a test image. Overall, the average model face was rated ‘male’ 83% of the time, and ‘female’ only 17% of the time, indicating a strong tendency of raters to rate the average face as ‘male’, χ^2 (1) = 249.93, $p < 0.0001$.

APPENDIX 5

Distances between face landmarks: A total of 14 distances based on 56 face landmarks previously described in Chakravarty et al. (2011).

#	Landmark distances
15-16	Left eye length
11-10	Right eye length
17-18	Left eye height
12-13	Right eye height
4-9	Forehead height
22-2	Face width at eyebrows
42-43	Face width ear to ear
54-56	Face width cheekbone
25-26	Nose width
29-35	Mouth width
22-21	Nose length
20-32	Nose to upper lip
20-44	Nose to tip of chin
37-44	Bottom lip to tip of chin

APPENDIX 6

Distances between skull landmarks (15)

#	Landmark distances
1-3	Distance between the inner corners of the eye sockets
1-2	Length of the left eye socket
3-4	Length of the right eye socket
6-7	Distance between mandibular sinuses
6-9	Size of the left mandibular sinus
7-8	Size of the right mandibular sinus
12-13	Right-third-molar to left-third-molar
19-12	Tip of chin to left-third-molar
19-13	Tip of chin to right-third-molar
10-11	Spine to lower front-teeth
10-19	Spine to tip of chin
19-11	Tip of chin to lower front-teeth
5-19	Nose bridge to tip of chin
5-11	Nose bridge to lower front-teeth
5-14	Nose bridge to upper front-teeth

APPENDIX 7

Relationship between the face PCs described in Chapter 3 (current) and face PCs described in Chapter 2 (Mareckova et al., 2011). Significant correlations greater than 0.4 are listed.

Face PC (current)	Face PC (Mareckova et al., 2011)	Pearson's r	p-value
1	1	0.86	$p < 0.001$
2	3	-0.86	$p < 0.001$
3	2	-0.82	$p < 0.001$
4	4	0.67	$p < 0.001$
5	6	0.56	$p < 0.001$
7	6	-0.45	$p < 0.001$
7	9	-0.44	$p < 0.001$
10	10	-0.42	$p < 0.001$
9	8	0.40	$p < 0.001$

APPENDIX 8

Effect of menstrual cycle and oral contraception use on amygdala BOLD response to faces

Adult women from Experiment I

Similar to Tahmasebi et al. (2012), masks for right and left amygdala were created separately for ambiguous and angry condition in the following way: First, we created thresholded images of *z*-statistics ($z>2.3$) for the following four groups: (1) OC during menstruation; (2) OC during mid-cycle; (3) freely cycling during menstruation; and (4) freely cycling during mid-cycle. Second, an intersection with Tahmasebi et al.'s (2012) amygdala mask was created for each of these four images. Third, a union of these four intersections was created. This was repeated separately for the ambiguous and angry contrasts, respectively. Unionized masks for these two conditions are presented in Figure 1A. Mean BOLD response in these unionized masks was calculated by Featquery (FSL 4.1.1) and its relationship with pill status and cycle phase was assessed.

Table 4A shows amygdala BOLD response (means, SDs) for each of the four groups: freely cycling menstruation, freely cycling mid-cycle, pill menstruation, pill mid-cycle. A two-way repeated measures ANOVA showed no significant effect of contraceptive Pill, cycle Phase, or their interaction on the mean BOLD response in the amygdala (Table 4B).

Table 4A: Amygdala BOLD response during menstruation and mid-cycle of pill and no pill group.

Mask	Freely cycling, Menstruation	Freely cycling, Mid-cycle	Pill, Menstruation	Pill, Mid-cycle
Right amygdala Ambiguous condition	M=0.06 SD=0.15	M=0.08 SD=0.13	M=0.09 SD=0.14	M=0.2 SD=0.13
Right amygdala, Angry condition	M=0.08 SD=0.1	M=0.11 SD=0.15	M=0.05 SD=0.12	M=0.12 SD=0.15
Left amygdala, Ambiguous condition	M=0.03 SD=0.09	M=0.08 SD=0.13	M=0.09 SD=0.11	M=0.12 SD=0.09
Left amygdala, Angry condition	M=0.06 SD=0.1	M=0.13 SD=0.13	M=0.1 SD=0.12	M=0.09 SD=0.09

Table 4B: Effect of Pill and Phase on amygdala BOLD response as assessed with two-way repeated measures ANOVA.

Mask	Effect of Pill	Effect of Phase	Pill * Phase interaction
Right amygdala, Ambiguous condition	F=2.73, p=0.11	F=2.26, p=0.14	F=1.16, p=0.29
Right amygdala, Angry condition	F=0.01, p=0.91	F=1.4, p=0.24	F=0.16, p=0.69
Left amygdala, Ambiguous condition	F=2.18, p=0.15	F=1.59, p=0.22	F=0.05, p=0.82
Left amygdala, Angry condition	F=0.00, p=0.99	F=0.55, p=0.46	F=1.35, p=0.25

Adolescent girls from Experiment II

Masks for right and left amygdala were created separately for ambiguous and angry condition in the following way: First, we created thresholded images of *z*-statistics (*z*>2.3) for the following two groups: (1) OC and (2) freely cycling girls (n=55 for each group). Second, an intersection with Tahmasebi et al.'s (2012) amygdala mask was created for each of these two images. Third, a union of these two intersections was created. This was repeated separately for the ambiguous and angry contrasts, respectively (Figure 4B). Unionized masks for these two conditions are presented in Figure 1B. Mean BOLD response in these unionized masks was calculated by Featquery (FSL 4.1.1).

Table 4C shows amygdala BOLD response (means, SDs) in the pill and no pill group. Left amygdala showed a trend for a lower mean BOLD response in the pill group (vs. no pill group) during the angry condition ($t_{(109)}$=-1.83, p=0.069). Mean BOLD response in the left amygdala during ambiguous condition and right amygdala during both ambiguous and angry condition did not show any significant difference when comparing adolescent girls in the pill vs. no pill group (left FFA ambiguous: $t_{(109)}$=-1.25, p=0.21; right FFA ambiguous: $t_{(109)}$=-1.46, p=0.15; right FFA angry: $t_{(109)}$=-1.48, p=0.14).

Table 4C: Amygdala BOLD response in pill and no pill group.

Mask	**Freely cycling**	**Pill**
Right amygdala, Ambiguous condition	M=0.19 SD=0.15	M=0.15 SD=0.12
Right amygdala, Angry condition	M=0.16 SD=0.16	M=0.12 SD=0.17
Left amygdala, Ambiguous condition	M=0.14 SD=0.14	M=0.11 SD=0.012
Left amygdala, Angry condition	M=0.13 SD=0.15	M=0.08 SD=0.14

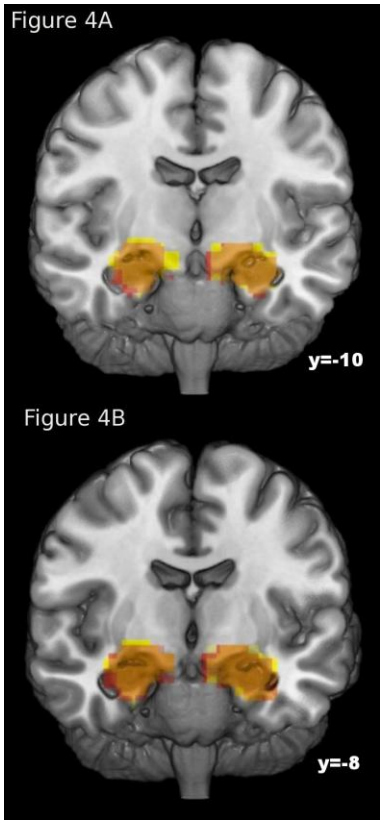


Figure 4: Unionized amygdala masks created for angry (red) and ambiguous (yellow) condition in Experiment I (Figure 4A) and Experiment II (Figure 4B). Intersection of these masks is displayed in orange.