The Dissertation Committee for Philip Matthew Bronstad Certifies that this is the approved version of the following dissertation:

Theoretical and Methodological Congruence with Face Perception Research: An Alternate Paradigm for Facial Attractiveness

Committee:

____
Judith H. Langlois, Supervisor

____
Lawrence K. Cormack

____
Randy L. Diehl

____
Patrick K. Randall

____
Michael J. Ryan

Theoretical and Methodological Congruence with Face Perception Research: An Alternate Paradigm for Facial Attractiveness

by

Philip Matthew Bronstad, B.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August, 2004

Dedication

To my family.

Acknowledgements

Most of all, I want to thank Judy Langlois with whom I was very fortunate to work. She is supportive of all of her students, offering wisdom, intelligence, partnership, and financial support. We have a great lab, with current and former students Angela Griffith, Becky Hoss, Jennifer Ramsey, Lisa Rosen, and Teresa Taylor-Partridge. I value them all as friends and colleagues. I should also thank the many undergraduates who worked in the lab and who made this dissertation possible, including Aakash Desai, Briana McKeague, James Norris, and Roberto Rojas.

Attending a large school like UT Austin enables graduate students to assemble top-notch dissertation committees. Each of my committee members helped me in an important way. Larry Cormack and his students helped me better understand modeling; Randy Diehl greatly influenced the tone of the dissertation; Pat Randall consulted with me for years and suggested using partial least squares; Mike Ryan filled in many of the gaps in my knowledge of sexual selection.

There are many people who I was fortunate to meet during graduate school who shaped the experience and this document. John Slatin and Kay Lewis, at UT's Accessibility Institute, are two colleagues with whom I had fruitful intellectual partnerships. Although we worked on subject matter unrelated to this dissertation, they encouraged me to work on problems that were often analogies of problems dealt with in this dissertation. Ted Barker, Mike Bergman, Nate Marti and others at ITS statistical and mathematical consulting were part of a subculture (of which I was a proud part) of statistics geeks and who offered advice and criticism. Neill Hadder helped me to understand the anthropologists' point of view. Other people whose comments and help I appreciated include Richard Russell, Umesh Rajashekar, Alice O'Toole, Risto Miikkulainen, and Tal Tversky.

Last, thanks and love to my wonderful family, my parents Walt and Pat Bronstad, my sisters Kris and Kate Bronstad, and those who are practically part of the family, Peggy and Hannah Whilde. They all supported and encouraged me greatly. My wife, Cheryl Browne, kept me going in graduate school. I turn to her for advice frequently and love her dearly.

Theoretical and Methodological Congruence with Face Perception Research: An Alternate Paradigm for Facial Attractiveness

Publication No._____________

Philip Matthew Bronstad, Ph.D.
The University of Texas at Austin, 2004

Supervisor: Judith H. Langlois

This dissertation critiques the prevalent and contemporary explanatory framework for facial attractiveness hypotheses that are based on certain assumptions of the relationships among hormones, facial growth, and immune system function. I propose an alternative explanatory framework based on face perception research. The facial attractiveness and facial perception literatures currently are not integrated; however, much of our knowledge about face recognition is relevant to understanding how people make judgments of facial attractiveness. In particular, computational methods used in face recognition are vital to testing competing facial attractiveness hypotheses.

Three studies that test the two major hypotheses proposed to explain facial attractiveness, averageness and sexual dimorphism, are presented. Each study was designed to provide critical tests of these hypotheses as well as demonstrate how face representation models can be used for this purpose. Results show that both averageness and sexual dimorphism are correct, explaining different aspects of facial variation that covary with attractiveness judgments. Modeling results show that facial averageness

should be construed as the degree of similarity between a face and a hypothetical gender-neutral prototype rather than a sex-specific prototype.

Finally, this research demonstrates that unsupervised learning algorithms (principal components analysis and independent components analysis) can explain moderate amounts of variance in attractiveness. A supervised connectionist model, however, can explain all of the variance between faces in mean attractiveness ratings, generalizing almost perfectly to predict attractiveness judgments made to novel images of faces.

Table of Contents

List of Tables

List of Figures

Chapter 1: *Introduction*

“Physical attractiveness is, in many ways, a homely variable,” Berscheid and Walster wrote three decades ago in their authoritative review of the attractiveness literature (1974, p. 206). Despite its homeliness, it “demands respect” (p. 206) because of its intrusion into everyday life. Attractiveness research has since grown and demanded respect from scientific psychology.

Historically there are two conceptual schemes for beauty, objectivism and subjectivism. These approaches have deep roots and are diametrically opposed. In the objectivist approach, beauty exists independently of perceivers; it is a tangible quality of objects (that is, in the form). Plato’s aesthetics held that objects in the real world approximate ideals that exist in another world of pure forms to which we do not have access. Religious Neo-Platonism is another manifestation of the objectivist tradition, which holds that beauty is a reflection of divine influence.

In the second conceptual scheme, subjectivism, beauty does not exist in the world; it is a judgment that occurs in the beholder’s mind. According to Hume there can be no wrong judgment of beauty: “a thousand different sentiments, excited by the same object, are all right: Because no sentiment reflects what is really in the object.” (Hume, 1985, p. 230). Most nonscientists probably believe in subjectivist notions of beauty (Langlois et al., 2000).

Despite the prevalence of subjectivism among laypeople, the current conception of attractiveness within psychology is now predominantly objectivist. Lindzey (1965), as well as Berscheid and Walster (1974), successfully predicted that the study of facial morphology, would gain acceptance in attractiveness research. Lindzey proposed that morphology had been neglected because both its connection to phrenology pseudosciences and the presence of a strong Protestant ethic in American culture - the ethic that an individual achieves through hard work precludes any suggestion of influence by genetic inheritance. Morphology is the primary focus of contemporary facial attractiveness research, and the view that attractiveness can be related to the physical structure of faces certainly contributed to the belief that attractiveness is a property of faces.

Despite this, attractiveness research has become sidetracked. Many researchers have come to endorse a particular variant of objectivism[1]: the face is a set of symbols or cues that convey detailed information about genetic quality to potential mates. These researchers conceive of the feeling of attraction as a reflection of evolved mechanisms that evaluate these symbols. For example, Cunningham has stated that "…attractiveness may be demystified if a pretty face is merely seen as a symbol for desirable internal qualities." (1995, p. 277). Similarly, Thornhill conceives of beauty as "…the perception of cues to high reproductive potential…" (1998, p. 564).

After social psychologists overturned traditional views of beauty, discovering that people have similar notions of attractiveness, facial attractiveness research changed. Beauty proved not to be "in the eye of the beholder" as many people assumed. The consistency of the findings encouraged the study of attractiveness as a natural extension of the search for genetically-specified psychological mechanisms shaped by natural selection (Reis & Zeidel, 2001; Rhodes et al., 2001; Thornhill & Gangestad, 1993). Thirty years ago Eleanor Gibson wrote that psychologists do not believe "in innate ideas" (Gibson, 1969, p. 20), however, many psychologists who favor the evolutionary theories appear to believe that preferences for faces are largely genetically specified and present from birth (e.g., Etcoff, 1999; Fink, Grammer, & Thornhill, 2001).

[1] Penton-Voak and Perrett (2000) refer to a similar concept as *structuralism*, acknowledging that the point of view is predominant in Darwinian approaches to attractiveness.

My central thesis is that there are resolvable conceptual problems with contemporary facial attractiveness. Fundamentally, many researchers analyze faces and do not place enough emphasis on the perceivers of the faces. I will point out the problems with the state of the art, and indicate an alternate framework that is likely to be more productive. The alternate framework is not new; it is compatible with facial perception research and is, or appears to be, endorsed by some attractiveness researchers (e.g., Enquist, Ghirlanda, Lundqvist, & Wachtmeister, 2002; Rubenstein, Langlois, & Roggman, 2002). This dissertation's first important contribution is its contextualization of the dominant facial attractiveness research paradigm. In contrast, many contemporary overviews of facial attractiveness theories and research have adopted a nativist viewpoint (e.g., Fink & Penton-Voak, 2002; Thornhill & Gangestad, 1999). Second, my proposed tests of facial attractiveness hypotheses are created to be consistent with current knowledge of, and methods used in, face perception.

The importance of studying attractiveness

Attraction is mysterious and is often a thrilling feeling. Personal relationships are supremely important to people, so understanding what factors affect their formation and course is substantive. Physical attractiveness certainly is a strong determinant of whether attraction between two people blooms.

Moreover, attractiveness research influences how we see ourselves as a species. One view is that attractiveness is a reflection of mechanisms that helped our ancestors select a mate and continue their genetic lineage. A dichotomous alternative is that our conceptions of attractiveness reflect cultural influences that we readily emulate. Perhaps cognition and culture intersect; does perceptual learning subtly influence our notions of attractiveness? If so, what is the nature of the mechanism(s) in which biology and culture integrate (cf. Symons, 1995)?

Attractiveness research also may have large-scale consequences to society. For example, claims about which facial features cause faces to be attractive are frequently reported in popular science magazines and television. This information is useful to the $7 billion American cosmetic surgery industry, which actively commodifies beauty and uses multiple means to convince consumers of its importance. Unilever Corporation, a $47 billion multinational that spends more on advertising than all but two companies *worldwide* (Wentz, 2002), promotes its beauty products such as Dove™ soap and SlimFast™ diet shakes. Unilever is part of the industry that obviously recognizes, and capitalizes on, the value of this research. They have funded Perrett's research group - major proponents of one of the attractiveness hypotheses - since at least 1998 (Perrett, et al., 1998; Penton-Voak, Perrett, & Pierce, 1999).

More importantly, researchers can interpret results to deliberately influence the behavior of those in the medical professions. Thornhill and Møller hypothesize that attractiveness judgments are fundamentally detections of imperfections in appearance caused by "mistakes" during growth. They directly targeted medical professionals, publishing in *Biological Reviews* a paper containing the following statement: "We offer this paper as a 'wake-up call' to the health professions on the importance of developmental stability [i.e. attractiveness] as a marker of good health" (1997, p. 498). Thornhill has written elsewhere that attractiveness may be an indicator of health and "genetic fitness" (e.g., Thornhill & Gangestad, 1999); thus, if true, health workers should make decisions about treatment and diagnosis based on attractiveness, because certain patients -- *unattractive* patients -- are more likely to have health problems. The link between health and attractiveness, however, is tenuous (Langlois et al., 2000; Shackelford & Larsen, 1999), perhaps nonexistent or in the "wrong" direction (Kalick, Zebrowitz, Langlois, & Johnson 1998). Research in facial attractiveness can have serious economic and social consequences, especially given that our health care system faces several crises (Pellicer & Burke, 2002).

How Attractiveness is Measured

Berscheid and Walster called attractiveness a homely variable, in part, for its elusiveness. Attraction - the feeling that one person develops for another - is often idiosyncratic and mercurial. However, this is not the sense of attraction that contemporary facial attractiveness researchers try to explain. Rather, facial attractiveness research attempts to account for why initial impressions of facial attractiveness are so consistent from person to person. This sense of attraction is sometimes referred to as initial impressions at zero acquaintance (Zebrowitz & Collins, 1997).

This sense of physical attractiveness caught the attention of researchers because the degree of agreement among raters was initially very surprising. For example, Iliffe (1960) found that attractiveness judgments, made to images of 12 women, were correlated between .8 and .98 among groups of British participants. Despite variance in sex, age, social and class background, and geographic region agreement was very high between participant groups.

Attractiveness judgments often have high interrater agreement, even among people from different generations and cultures. Within-culture agreement is often very high, with effective reliabilities estimated to be +.9 (Langlois et al., 2000). Langlois et al. observed cross ethnic judgments to be correlated .54 and cross cultural judgments correlated .71. Good reviews can be found in Berscheid & Walster (1974), Langlois et al., (2000), and Shepherd, (1981). Given that there is a high degree of agreement within-cultures, virtually all researchers measure attractiveness on a unidimensional scale.

It should be mentioned that although some researchers have concluded that cross-cultural differences are inconsequential or nonexistent (for example, Fink, Grammer, &

Thornhill, 2001), this is not the case. Recorded between-culture agreement varies (e.g., Jones, 1996; Langlois et al., 2000), as do cosmetic practices. There are striking differences between how cultures alter their faces and bodies. For example, many preindustrial tribes institutionalized scarification and tattooing of the face, sharpening or blackening of teeth, or mutilation on or near the face to wear lip discs, bars in the nasal septum, et cetera (e.g., Adriani & Kruyt, 1951; Bruce & Young, 1998; Schultze, 1907). Nevertheless, there are usually positive correlations for preferences between cultures, indicating a general agreement among different cultures that is less strong than within-culture agreement.

INTRODUCTION TO THE MAJOR THEORIES AND PARADIGMS

Cognitive theories of facial attractiveness

Two psychological fields have influenced research on facial attractiveness: evolutionary psychology and, to a lesser degree, cognitive psychology. Langlois & colleagues (e.g., Langlois & Roggman, 1990; Rubenstein, Langlois, & Roggman, 2002) developed a facial attractiveness theory that is primarily cognitive, rather than evolutionary; averageness theory. Averageness theory holds that faces closer to the central tendency (or prototype) are more attractive than faces father from the prototype. It is called averageness, however, because researchers have often created simulations of face prototypes by pixel averaging images of faces (see Figure 1, below). Averageness is the sole facial attractiveness theory consistent with face perception research, including theories of representation, categorical learning, and research on the development of face expertise (Rubenstein, Kalakanis, & Langlois, 1999; Valentine & Bruce, 1986a).

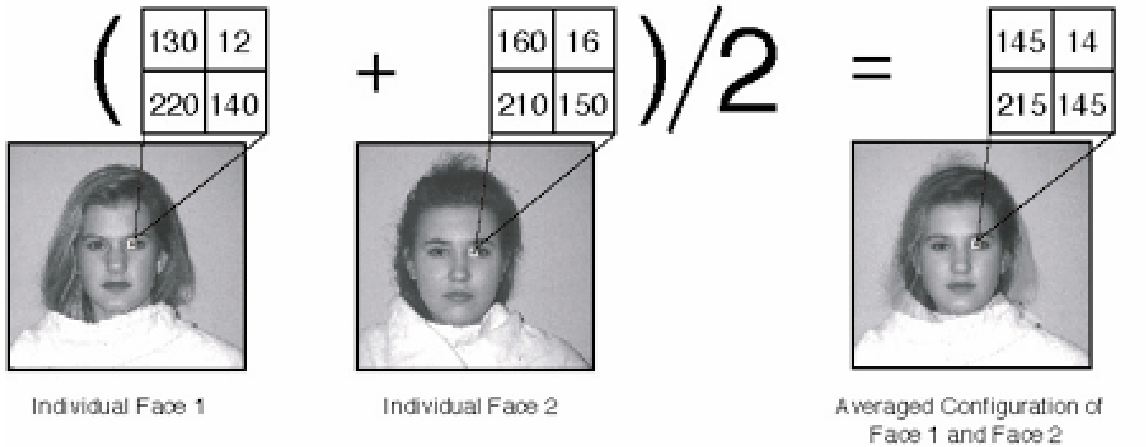

Figure 1: Pixel averaging to produce average faces (from Langlois, Roggman, & Musselman, 1994, used with permission).

A clear strength of the averageness hypothesis is its compatibility with the face perception literature. Averageness has a specific definition that relates to face structure, but it is also a theoretical component of a face recognition system, a prototype (Turk & Pentland, 1991; Valentine & Bruce, 1986b). Many formal models of face perception systems assume a facial prototype, as the central tendency is a convenient reference point for all other faces.

More formally, proponents frame averageness in terms of prototype theories of cognition (Langlois & Roggman, 1990). A prototype hypothesis of face perception maintains that individuals abstract the central tendency of the category of stimuli, and that faces are encoded in terms of how they uniquely deviate from the prototype (Busey, 2001). I use the word "prototype" neither to support nor test the prototype or exemplar hypotheses of knowledge representation but instead as a convenience and acknowledgment of the similarity between the presumed manner in which individual prototypes develop and how facial image composite are constructed.

The principal reason averageness theory's framework differs from other theories of facial attractiveness is that its major proponents are developmental psychologists. Langlois was the first to discover that young infants prefer to look at attractive faces than at unattractive faces (Langlois et al., 1987). Langlois's findings have often been interpreted as indicating an innate mechanism for facial attractiveness (e.g., Cunningham, Roberts, Barbee, Druen, & Wu, 1995; Etcoff, 1999; Perrett et al., 2002; Rhodes et al., 2001); however, subsequent investigation revealed that newborns do not exhibit preferences for attractive faces (Kalakanis, 1997). A reasonable explanation for the demonstration that newborns do not know which faces adults find attractive is that their face perception skills are undeveloped.

The ability of infants to discriminate faces develops through experience. Many researchers conceive of face perception as a learned skill (Gauthier & Nelson, 2001; O'Toole, Abdi, Deffenbacher, & Valentin, 1995; Stevenage, 1995), and developmental studies indicate that adults are better at many tasks than children, such as recognition and gender discrimination (Johnston & Ellis, 1995). Newborns prefer to look at faces than other comparably complex stimuli, however, suggesting an early capacity for face detection that provides for rapid learning of variation among faces. This tendency to attend to face-like stimuli may be highly rudimentary in its specification (Cassia, Turati, & Simion, 2004).

Even in the absence of developmental data, studies on adults indicate face perception is a learned skill. The "other-race" effect is the phenomenon that observers are better at remembering recently-learned faces if the individuals pictured are of the same race as the one with which the observer has experience (Valentine, Chiroro, & Dixon, 1995). Such investigations into the causes of observations such as "they all look the same to me," suggest that some degree of expertise with local racial or cultural variation in faces is needed to recognize individuals. Second, although adult face perception is sophisticated, people perform poorly on recognition tasks with inverted faces (Valentine, 1988). The "Margaret Thatcher" illusion is a good demonstration of how facial inversion disrupts face processing (Thompson, 1980). In Thompson's illusion, the eyes and mouth of a face are inverted, making the face appear grotesque. When viewed upside-down, however, the face appears normal. That we can be "tricked" in this manner indicates that we are experts at perceiving *upright* faces.

Figure 2: George Bush, "Thatcherized."

Evolutionary psychological theories of facial attractiveness

The dominant evolutionary facial attractiveness paradigm is the immuno-endocrinological theory of facial attractiveness. It is derived from evolutionary psychological theory, and is also called a "good-gene" theory of sexual selection. Immuno-endocrinological theory holds that the purpose of attractiveness mechanisms is to help people select mates who are healthy and whose genes will protect offspring from disease.

Two evolutionary psychological hypotheses about the properties of faces that determine their attractiveness are symmetry and sexual dimorphism. The symmetry hypothesis proposes that faces more symmetrical are more attractive than less symmetrical faces (Grammer & Thornhill, 1994). The sexual dimorphism hypothesis proposes that the properties of faces that make them appear masculine or feminine are also the properties that make them appear attractive or not (Cunningham, 1986; Perrett et al., 1998).

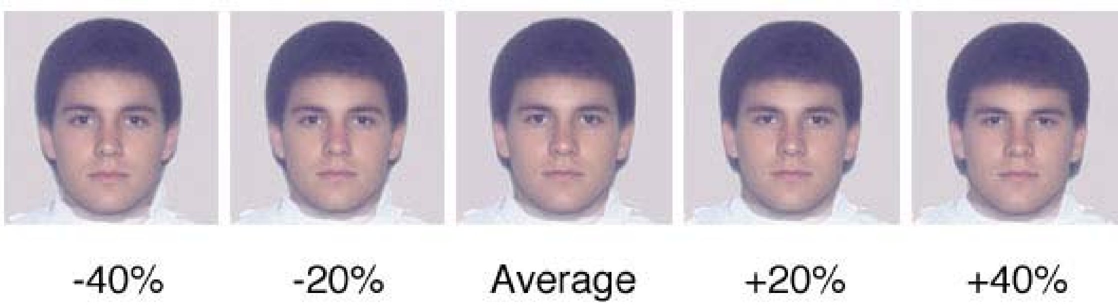

Figure 3: Images manipulated to vary in facial masculinity (after Perrett, et al., 1998).

In the last twenty years, much facial attractiveness research has been generated. In normal scientific progress competing theories become eliminated, rather than accumulate. Bruce and Young (1998) observed that, although it shows promise, the field is still in the initial stages of development. Surveying the field, Rhodes and Zebrowitz found that there is no “gold standard of facial attractiveness” (2002, p. viii), noting diplomatically that all of the different hypotheses about what make faces attractive seem to be equally important.

The assessed lack of progress in facial attractiveness research may be due to the perspective commonly applied to the problem. In evolutionary psychology (EP), researchers analyze the stimulus rather than the perceiver, focusing on what faces might reveal about the person being observed. For example, a common method is to measure the distances between facial landmarks and correlate the feature distances with attractiveness ratings (e.g., Cunningham, 1986; Grammer, Fink, Juette, Ronzal, & Thornhill, 2002; Penton-Voak, et al., 2001), suggesting that attractiveness is a property of faces and that face perception (e.g., detection, recognition) is not relevant.

Much of face attractiveness methodology is guided by assumptions about how sex hormones affect facial growth rather than how the visual system processes faces, or how knowledge of faces guides perception of features. Moreover, the facial measurement methods used in attractiveness research are very different from methods used in research on face recognition and representation.

There is a tendency for evolutionary psychological hypotheses of attractiveness perception to be “black box” models. For example, Tooby and Cosmides (1992) state that it is unimportant to understand physiological mechanisms in detail, and what is crucial is to understand to which stimuli organisms attend and how they then behave. “Knowledge of this hardware, however, is not necessary for understanding the programs as information-processing systems” (Tooby & Cosmides, 1992, p. 66). One reason that physiological mechanisms are not well-studied is that analysis may end up “bogged in a vast intricacy of unrelated detail" (Russell, 1945, quoted in Tinbergen, 1976, p. 152).

In another sense in which EP attractiveness hypotheses are black box mechanisms, it is not always apparent that the cues to good genes are perceivable. Scheib, Gangestad, and Thornhill (1999) reported that in their study of symmetry and facial attractiveness that, despite a moderate correlation between measured symmetry and perceived attractiveness, participants’ judgments of facial symmetry did not correlate with measured symmetry. Swaddle & Ruff (2004) wished to answer the question of whether small deviations from perfect symmetry, such as those found naturally in animals and are supposedly a cue to good genes, are even perceived by potential mates.

They found that European starlings can not detect small deviations from perfect symmetry. To put it naïvely, if cues to good genes are not perceivable, then how do organisms select mates who have good genes? The answer is that we must try to understand how animals and humans perceive others. Whether their preferences are adaptive is a separate question. Swaddle and Ruff stated that the results “help us to focus on traits that are relevant to the ways in which birds see their world” (p. 38). I will argue the same point for the study of human facial attractiveness.

It is necessary to lay bare the theoretical elements and structure common to Evolutionary Psychological facial attractiveness theories. In this section, I will show 1) the scope of EP facial attractiveness theory; 2) the actual scope of EP facial attractiveness research is a subset of the overall theoretical scope, and; 3) the elements of EP attractiveness theory outside the scope of EP research are unsupported by research in the relevant fields. I will further show that the methods employed by EP do not complement research into human face perception. These steps are necessary to establish what types of evidence constitute tests of facial attractiveness hypotheses and what types of evidence do not.

Are we perceivers of genetic quality or perceivers of faces? This appears to be a false dichotomy, but the question of what we perceive is relevant given the discourse of the attractiveness literature. For example, some EP researchers prefer to talk of face features not as noses, chins, etc., but as *hormone markers* (Fink & Penton-Voak, 2002; Grammer, Fink, Juette, Ronzal, & Thornhill, 2002; Thornhill & Møller, 1997). For example, Johnston et al. recently concluded that covariation of attractiveness ratings and certain stimulus properties was "convincing evidence that participants' choices were strongly influenced by hormone markers" (2001, p. 263). The choice of vocabulary (“hormone marker”) places the meaning of the results into the realm of biological systems. Is psychological experience less meaningful than biological reality? Alternatively, Johnson et al. may have been suggesting that hormones and biological mechanisms are the immediate causes of our preferences. Although biological mechanisms must underlie attractiveness preferences, it is not known how hormone action causes changes in facial structure. It is questionable that faces have hormone markers and that it is questionable that humans perceive subtle variation in genetic quality by others' faces.

Thornhill and Møller suggest that hormones, facial appearance, and parasites are meaningfully linked:

> "High-quality males will be able to develop large sex traits, cope with high levels of androgens, and only compromise their immune defense to a relatively small extent. Sex differences in the course of parasite infections and relationships among sex hormones and parasitism are consistent with the immunocompetence handicap hypothesis." (Thornhill & Møller, 1997, p. 504).

The immuno-endocrinological hypothesis structure, its common elements and relationships, are outlined below (Figure 4).

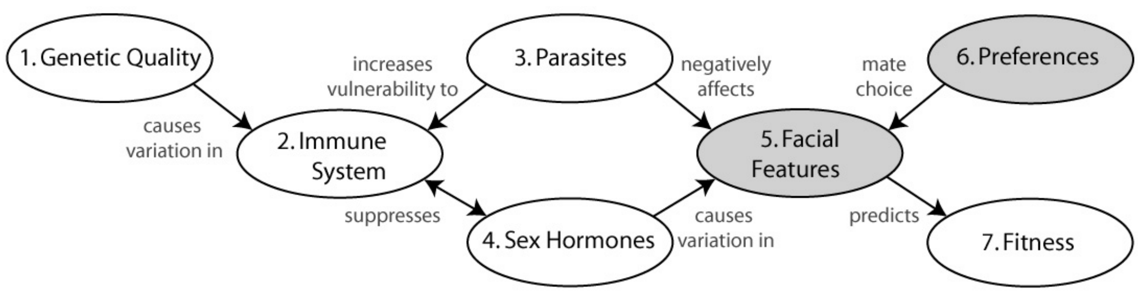

Figure 4: The immuno-endocrinological hypothesis of facial attractiveness.

Not shown in Figure 4 is the feedback process by which preferences are changed by natural selection. Regardless, this representation describes existing EP facial attractiveness hypotheses. For example, we can describe Thornhill's theory of developmental stability by referring to the numbered boxes in Figure 4: (1) heritable variation in genetic quality of individuals that (2) makes some individuals more susceptible to (3) environmental pathogens such as rapidly-evolving parasites. Further, as individuals develop, their ability to defend against such pathogens is dependent on their output of (4) sex hormones, especially during puberty. Specifically, production of sex hormones is hypothesized to inhibit immune system function. Thornhill & Møller (1997) hypothesize that only individuals of high genetic quality can afford the hit to their immune system that high levels of sex hormones generate. EP researchers propose that testosterone promotes facial growth whereas estrogen inhibits growth (Symons, 1995; Thornhill & Grammer, 1999). Further, (5) between-individual variation in facial features is linked to sex hormone output *and* is affected negatively by pathogens such that individuals can identify and select mates with good genes on the basis of their facial features. These preferences, given a genetically-transmittable basis, will be passed onto offspring.

EP researchers often argue for the validity of good genes models and their applicability to humans by analogy – pointing to the extreme features of non-human animals, such as large antlers, thought to be sexually selected (e.g., Cunningham, Druen, & Barbee 1997; Thornhill & Gangestad, 1999; Penton-Voak, et al., 2001). Simulation results (Kirkpatrick, 1996) suggest that good genes models can function as hypothesized, but that the indirect benefits that females receive by choosing mates that have good genes can not outweigh any the selective advantage of direct benefits they can obtain from mates. For example, males of some species males give a courtship "gift" that the female consumes to increase her fertility. In other species, males provide paternal care. More problematic for good genes models is the lack of evidence that females can increase their reproductive success by mating with males that are free from parasites (Ryan, 1997).

Despite the lack of evidence, the logic of the immuno-endocrinological theoretical structure is understandable and seems workable. Despite the intuitiveness of the model each theoretical element needs to be tested and shown to be plausible. In Figure 4, the boxes that are colored gray indicate the scope of EP facial attractiveness research. The way in which EP theories have been tested is to look at the association between features and attractiveness[2] (Figure 4, boxes #5 and #6). Showing that variation in facial appearance relates to variation in attractiveness can falsify specific hypotheses of facial attractiveness, but this kind of data cannot be accepted as evidence for the rest of the theoretical structure. A problem with testing EP attractiveness hypotheses is that the hypotheses depend on many theoretical elements and relationships among the elements that are not usually tested.

Some of the connections are plausible whereas others are not. In particular, it is plausible that parasites affect both facial features and the immune system. Burkitt's lymphoma, which may be facilitated by malarial infection, is an example of a tropical illness that dramatically affects the facial appearance of infected individuals. Children and adolescents mainly develop the lymphoma, which can cause massive jaw tumors (Palmer & Reeder, 2001). Moreover, facial appearance and preferences are plausibly linked. As noted already, people have fairly consistent preferences for attractive and healthy individuals (Langlois et al., 2000).

Less plausible are several of the remaining connections between EP's theoretical elements shown in Figure 4. First, is immune function heritable? The immuno-

[2] Although studies have shown that the connection between feature (box #5 in Figure 4) and fitness (box #7) does not exist (e.g., Kalick, Zebrowitz, Langlois, & Johnson 1998), EP researchers correctly argue that this link does not provide a test of whether attractiveness preferences are evolved modules (Thornhill & Gangestad, 1999). Current-day fitness is not expected to conform to the relationship(s) of feature and fitness when preferences evolved. Thornhill & Gangestad state that hypothesis tests carried out in preindustrial tribes would be valid. The argument against testing the hypothesis in industrialized societies, however, holds for preindustrial societies as well. Thus, the hypothesis is not testable.

endocrinological hypothesis requires that there be heritable variation in immune system function. Most of the evidence on this point comes from studies of the major histocompatibility complex (MHC) genes, which are involved in how the immune system distinguishes the body from external threats. Penn and Potts (1999) reviewed MHC studies to determine what adaptive role the MHC genes play. One possibility is that having heterozygous alleles at MHC loci helps individuals avoid disease, and so, individuals choose mates so that their offspring will be heterozygous. This is called the heterozygote advantage hypothesis. A second possibility is that individuals choose mates with different MHC alleles to avoid inbreeding depression. The two possibilities are not mutually exclusive. However, Penn and Potts showed that whereas there is evidence that mate choice to avoid inbreeding may occur (for example, Wedekind & Füri, 1995), there is little evidence to support the heterozygote advantage hypothesis. Penn and Potts suggest that the hypothesis has not been adequately tested, indicating that it might be supported by 1) tests of infestation by multiple, different, parasites or 2) studies of rapidly mutating parasites such as HIV. In any case, that the heterozygote advantage hypothesis has not received empirical support is problematic for the immuno-endocrinological hypothesis.

Second, the immuno-endocrinological hypothesis states that hormones, especially testosterone, suppress the immune system. It is commonly accepted that hormones suppress the immune system, however, Fink & Penton-Voak (2002) recently pointed out that the connection between facial features and immuno-competence has received no empirical support. Immuno-competence could have an indirect effect on facial features. Studies of immune system response, however, have not supported the immunosuppression hypothesis that normal variation in hormones negatively and globally affects the immune system (see reviews in Hasselquist, Marsh, Sherman, &

Wingfield, 1999; Hillgarth & Wingfield, 1997). Braude, Tang-Martinez, and Taylor (1999) hypothesized that the assumed immunosuppressive effects of testosterone have not been found because, rather than suppressing immunity, testosterone may instead affect the distribution of the body's immune response. They suggest that, in a stress response, testosterone causes immune cells to migrate from the bloodstream to the skin, where the immune system is ready for response to injury. Such a redistribution could be mistaken by researchers for a global suppression of function. Injuries are common in male-male competition, so Braude, Tang-Martinez, and Taylor believe that the redistribution of leukocytes is an adaptive anticipatory response. Stress-induced immune function redistribution, mediated by corticosteroids, has been demonstrated (see review in Dhabhar, 1998). That immunosuppression is in question is problematic for the immuno-endocrinological hypothesis of facial attractiveness.

The connection between hormones and facial appearance seems to be one of the strongest links in the immuno-endocrinological theoretical structure. Testosterone is assumed to promote growth in dimorphic areas of the face, whereas estrogen is presumed to inhibit growth in those areas (Cunningham, Barbee, & Philhower, 2002; Grammer & Thornhill, 1994; Johnston, et al., 2001; Symons, 1995; Thornhill & Gangestad, 1997; Thornhill & Møller, 1997). In contrast to this assumption, the literature about facial appearance and hormones does not support the hypothesis. There is no evidence that facial features could reveal the type of information that EP facial attractiveness researchers assume they do. More specifically, within-sex variation in facial appearance, masculinity and femininity, has not been shown to relate to hormonal differences (Tanner, 1990). Additionally, between-sex differences in facial appearance are not attributable to testosterone and estrogen in the ways assumed by the immuno-endocrinological model (e.g., Grumbach, 2000). I will discuss this in a later section of this paper.

The immuno-endocrinological theoretical structure has an intuitive appeal and is by no means disproved here. The plausibility of the theory, however, is undermined by the apparent unsoundness of several of the theory's critical assumptions and sub-hypotheses. This does not preclude the possibility of an evolutionary explanation for facial attractiveness, nor that a secondary explanation that complements the current EP theory structure.

I next will discuss the major facial attractiveness hypotheses in more detail, discussing their theoretical underpinnings, reviewing research germane to each, and offering constructive critiques before detailing my proposed studies.

Chapter 2: *Facial Attractiveness Research*

SEXUALLY DIMORPHIC FEATURES

Several theories of facial attractiveness refer to the difference between men's and women's faces and additionally posit that the differences reflect distinct developmental processes (Johnston et al., 2001; Perrett et al., 1998; Thornhill & Møller, 1997). These theories reflect the general immuno-endocrinological model of EP facial attractiveness research, that sex hormones are hypothesized to cause sexual dimorphism in facial appearance, both between and within the sexes. Before puberty, boys and girls faces are distinguishable, but are extremely similar in form (Wild et al., 2000; Tanner, 1990). At puberty, the faces of boys and girls grow divergently; although they follow similar trajectories, men's faces grow more in several areas: primarily the sinus cavity and brows (making the forehead more pronounced), and the upper and lower jaws (Tanner, 1990). According to these theories, individuals must balance a trade-off between fighting parasites and becoming sexually dimorphic; because sex hormones are hypothesized to suppress the immune system (Cunningham, Barbee, & Philhower, 2002; Fink & Penton-Voak, 2002; Grammer et al., 2002).

These theories differ more in how they operationalize sexual differences in facial appearance than in theoretical foundation. Sexual dimorphisms can be described both in terms of local features, such as the nose or chin, and in terms of overall configuration of features. Local features are ill-defined, partly because the definitions used reflect the method of measurement researchers choose. It is often convenient to think of local features as being commonly-known parts of the face or computed directly from an image. Local features are sometimes called first-order features (Cottrell, Dailey, Padgett, & Adolphs 2001), some of which have names, such as eye, cheekbone, or hairline. Other features are distances between facial landmarks, such as the distance between the tip of the nose and the bottom of the chin. Cottrell et al. refer to these distances as second-order features, because the locations of the local features must first be determined before the distances between them can be computed. I will refer to them as *local* features, or distance measures, to distinguish them from *configural* features. We do not usually have names for configural features; however, we sometimes use adjectives such as “skinny,” “rugged,” or (un)attractive to describe configural differences. More complicated configural differences distinguish men's and women's faces (Burton, Bruce, & Dench, 1993; Perrett et al., 1998).

Three approaches to investigating hypotheses of sexual dimorphism exist. The first is the *Multiple Fitness Model* (MFM) a second-order feature hypothesis tested by taking feature measurements of real faces (Cunningham, Barbee, & Pike, 1990). The second I will refer to as *Composite Indexes*, in which feature distances are combined to give a single, configural, measure of facial masculinity (Scheib, Gangestad, & Thornhill, 1999; Penton-Voak et al., 2001). The third is a configural feature hypothesis tested by using computer morphing, a technique in which one face can be smoothly changed into another. Morphed variations between male and female faces are assumed to accurately represent sexual dimorphism in facial hormone markers, so this is generally called the *Hormone Marker* hypothesis (Perrett et al., 1998).

The Multiple Fitness Model

The MFM states that different facial features convey different types of information to the perceiver about the perceived. The perceiver tries to satisfy several different motivations in the search for a mate. Thus, attraction is an indication that the person observed fulfils these motivations: “Physical attractiveness may be demystified if a pretty face is merely seen as a symbol for desirable internal qualities” (Cunningham et al., 1995, p. 277). Cunningham defines five categories of symbols (features): 1) neonate; 2) mature; 3) expressive; 4) grooming, and; 5) senescence. Each feature category is named for what information it is hypothesized to convey. Cunningham does not specify how the different features are integrated by the perceiver, but maintains that “the whole may not be substantially greater than the sum of its parts” (Cunningham, 1986, p. 932). The MFM treats men’s and women’s faces differently; sexually mature women’s and sexually mature men’s faces are seen as dimorphic.

Cunningham’s early research preceded most of the evolutionary psychological facial attractiveness research, though the research programs are similar. Cunningham’s (1986) attractiveness assume sex differences in ideal (attractive) form and maintain that faces symbolically convey information about the quality of the person being observed. Thus, these investigations about which features individuals attend to begin with what researchers believe should be perceived; features that ought to be indicative of sex or reveal hormonal history. Such researchers advocate that people evaluate faces based upon those facial features that convey evolutionarily important qualities (i.e. “good genes”).

Cunningham (1986) defines features that symbolically convey personal attributes as linear distances between fiducial landmarks (i.e. easily identified parts of the face such as the pupil or bottom of the chin) (see Figure 5). For example, large eyes and small noses, indicated by distance between top and bottom of the visible eye and distance between the forehead bridge and the tip of the nose, respectively (each calculated from frontal face images), are theorized to:

> “…convey an exaggerated appearance of youthfulness, freshness, naiveté and openness… Adults who possess neonate features, such as large eyes or small noses, may elicit the attention and nurturance responses that evolved for youngsters" (Cunningham, Barbee, & Philhower, 2002, pp. 201 - 202).

Research on the Multiple Fitness Model

There is some evidence that linear facial feature measurements taken from photographs correlate with attractiveness ratings. Cunningham and others (e.g., Grammer, et al., 2002) analyze many feature distances to determine whether they predict attractiveness ratings. Certain measures correlate with attractiveness. Somewhat replicable effects include measures of eye size, cheek width, and eyebrow height in women (Cunningham, 1986; Cunningham et al., 1995; Grammer & Thornhill, 1994; Grammer et al., 2002) and chin or jaw width in men (Cunningham, Barbee, & Pike, 1990; Grammer & Thornhill, 1994). Obtained correlation strengths between feature values and attractiveness ratings are typically small. Cunningham showed that multiple linear regression (Cunningham, 1986; Cunningham et al., 1995) could explain substantial amounts of variance in attractiveness ratings (53% for a sample of 50 women). The analysis used eye height and nose area (neonate features), cheek width (maturity feature), and smile width (expressive feature) (Cunningham, 1986), but only the features that correlated most highly with attractiveness were selected for the model.

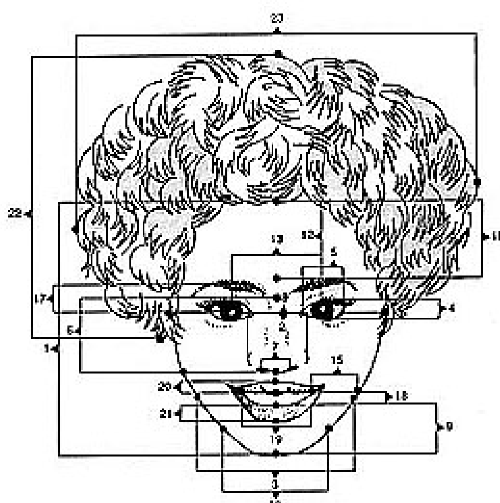

Figure 5: Cunningham's facialmetrics (from Cunningham, et al., 1995, used with permission).

Critique of Multiple Fitness Model studies

Cunningham advanced the study of attractiveness; his work certainly did a lot to suggest to researchers that attractiveness is measurable and determining how to predict what faces are attractive is a topic for study within psychology. There are, however, two important conceptual problems with the MFM. First, the most attractive faces must strongly reflect qualities of each category (i.e., mature, neonate, expressive, grooming, but not senescence), entailing that attractive faces are simultaneously neonate and sexually mature (Cunningham, Barbee, & Philhower, 2002). The decision of which features should be mature in form and which should be neonate in form seems arbitrary (c.f., Jones 1996). For example, why should the ideal woman's face have a neonate-like forehead but have a sexually mature lip? Why not neonate lips and a mature forehead?

The second conceptual issue is methodological. Researchers select any linear distances they wish to measure. We have little insight, however, into how researchers choose features to investigate. Faces are complex structures. There are thousands of possible feature distances that researchers can measure. For example, anthropologists and medical specialists use a set of skull landmarks that can be easily identified (Farkas, 1981). Farkas lists 43 landmarks that researchers can identify from frontal face images. Consider that for 43 landmarks there are 903 possible distance measures (ie., nonredundant pairs of the 43 landmarks), 12341 possible measurable angles (nonredundant triplets), and 407253 potential distance ratios (nonredundant pairs of the 903 distances). Which are relevant for attractiveness and which are not? Furthermore, Farkas's is a short list of the possible identifiable facial landmarks. For example, many researchers choose to record four locations around the visible part of the eye (e.g., Farkas, 1981; Thornhill & Grammar, 1994). There is no reason researchers could not determine the locations of four points around the iris, or eight around the eye, or the topmost part of the highest wrinkle on the forehead. Moreover, feature distances are not sufficient to account for face recognition or sex discrimination (Burton, Bruce, and Dench, 1993; Davies, Ellis & Shepherd, 1978)

Researchers report correlations between feature distances and attractiveness, but they usually investigate a small number of feature distances. Rarely reported, however, is how often linear-feature measurements correlate significantly with facial attractiveness. By chance five percent of randomly selected feature measures should correlate significantly with attractiveness if there is no real relationship between facialmetrics and attractiveness. Cunningham (1986) reported 21 feature measurement correlations for a sample of 50 women's faces and again for a subsample of only the 23 college seniors included in the full sample (the remainder were beauty pageant contestants). Although 6 of the measurements are significant and replicated in both samples (correlations with attractiveness ranging from $r = .29$ to .58), the average correlation with attractiveness rating is small, .16 for the full sample and .04 for the sample of college seniors.

Similarly, Grammer et al. (2002) reported 108 correlations of 36 features and attractiveness for 70 images of women (rated in 3 different poses), with separate predictions for each feature. The absolute value of the correlations between feature values and attractiveness ratings ranged from -.29 to .47. The average correlation, not reported, was close to zero: 0.07. This would seem to be a refutation of the utility of simple feature measures and call into question the practice of placing only the most highly correlated facial measurements into a multiple regression analysis (e.g., Cunningham, 1986, Cunningham et al., 1995).

Composite indexes

Other evolutionary psychologists testing sexual dimorphism hypotheses of facial attractiveness developed indexes of masculinity or femininity by combining some of

Cunningham, Barbee, & Pike's (1990) feature measurements (e.g., Penton-Voak et al., 2001; Scheib, Gangestad, & Thornhill, 1999). Using a composite index, rather than scattered feature measurements, is sensible. Researchers who have used composite index measures do not necessarily hypothesize that facial features reflect different internal qualities; instead, they propose a single-factor model - facial features jointly reflect how much and which hormones affected an individual's facial growth.

Research on composite indexes

Scheib, Gangestad, and Thornhill (1999) used a composite that was a simple combination of standardized measures of cheekbone prominence and lower face length, hypothesizing that as these features are sexually dimorphic (e.g., Tanner, 1990), a combination of distance measures that reflect variance in these features should be indicative of facial masculinity. Penton-Voak, et al., (2001) adopted Scheib et al.'s index, but added several feature distances not used by Scheib et al.. Penton-Voak et al. combined linear feature measures of eye size, lower face size, cheekbone prominence, face width, and eyebrow height.

Penton-Voak et al. (2001) showed that the feature distances used were sexually dimorphic but did not investigate whether the composite indexes correlated with judgments of masculinity. Penton-Voak et al. and Scheib, Gangestad, and Thornhill (1999) found that faces whose composite index measurements indicated the faces were masculine were more attractive than those measured as more feminine. Scheib, Gangestad, and Thornhill's composite index correlated .48 with attractiveness ratings of the men's faces. Attractiveness and Penton-Voak et al.'s composite index were correlated .26 and .21 with men's and women's judgments of men's facial attractiveness, respectively.

Critique of composite indexes studies

Although it is conceivable that individuals use distance between features when they discriminate men's and women's faces, there are no attempts in the face attractiveness literature to use statistical methods to combine features. In all previous examples, several features were chosen and given equal weight. One of the researchers reported whether that the measurements combined in the masculinity indexes were sexually dimorphic, but not how well they could be used in combination to predict whether a face is a man or woman (Penton-Voak et al., 2001).

Discriminant Function Analysis (DFA) is a statistical analysis that can be used to find optimal weights for a linear combination of features to classify faces as men or women. For example, Tanner (1990) reports how a simple composite measure, 3 x biacromial diameter (shoulder width) – 1x bi-iliac diameter (hip width), correctly classifies 90% of people by gender. In other words, DFA chose to exaggerate the prominence of men's shoulders by weighting (multiplying) shoulder diameter by 3 and negatively weighting the relatively smaller hip diameter to create a continuous measure on which a criterion can be placed. Below the criterion (small shoulders relative to hips), one predicts the person is a woman. Above the criterion one predicts the person is a man. Additionally, one could use split-half reliability to establish the efficacy of the feature combination - perform DFA on a set of men's and women's faces and then determine if the feature weightings predict the sex of a different set of faces. Burton, Bruce, and Dench (1993) used facial feature distances and DFA to discriminate images of men's and women's faces. They performed multiple, random, split-half reliabilities using many different facial feature measurements. Burton et al. used many 2-d distances, ratios and angles computed from the measurements, as well as 3-d distances derived from profile and facial photographs of the same person. Burton et al. found that using 12 distance measurements, each optimally weighted using DFA, gave 85% correct sex classification. People discriminate men's and women's faces relatively easily, however, performing at 96% correct classification for face images in which hair, jewelry, and makeup are not visible. Although the automated model seems close to human performance, Burton, Bruce, & Dench concluded that the most compelling message of their paper was the "sheer difficulty" of discriminating men's and women's face images using simple measures (p. 173). Analysis of the errors in sex classification led Burton et al. to conclude that the model used different classification rules than humans do. Burton et al. also constructed line drawings of the faces using the 2-D measures. In judging the sex of the line drawings, constructed with measures that their automated model used to correctly discriminate 85% of the faces by sex, human observers were only 59% accurate in judging the sex of the line drawings.

It is known that when famous faces are represented by veridical line drawings, human observers can recognize them only 47% of the time compared to 90% performance for the original images (Davies, Ellis & Shepherd, 1978). Further, photographic negation does not change relative placement of features but it impairs face recognition; participants recognized 55% of famous faces from photographic negatives whereas they recognized 95% of the same faces in photographic positives (Bruce & Langton, 1994).

Researchers have concluded that feature measurements are inadequate for automated face recognition systems and for modeling human face recognition (Bruce, Burton, & Hancock, 1995; O'Toole, Abdi, Deffenbacher, & Valentin 1995). "The primary problem with such codes is that they are often not adequate for quantifying and communicating enough information about an individual face to distinguish it from the multitude of competing similar candidates" (O'Toole, Wenger, & Townsend, 2001, p. 10).

As Burton, Bruce, and Dench's (1993) more complex and flexible feature distance system was inadequate (compared to humans) to discriminate men's and women's faces, simpler and rigid models (e.g., Cunningham, 1986; Grammer & Thornhill, 1994; Penton-Voak et al., 2001; Thornhill & Gangestad, 1994) are unlikely to adequately discriminate men and women's faces, or to predict perceptions of masculinity.

The hormone marker hypothesis

Perrett and his research group developed a different method of investigating how sexually dimorphic features relate to facial attractiveness (Perrett et al., 1998). They avoided many of the problems of attending to and combining smaller features by using a computer warping algorithm to treat sexual dimorphism as a set of configural differences. Warping is an image operation similar to morphing. Morphing is a nonlinear *fade* from one face to another – the configuration, color, and texture of face A changes gradually into that of face B. Warping is a variant of morphing in which the configuration, but not the texture and color, of face A changes into face B. Using computer software, a person manually creates a system of correspondence between the two faces, placing dots and curved lines at similar face locations, such as around the eyes, nose, mouth, and chin. Once the correspondence between two faces is established, the pixels in the image of one face can be transformed by the degree of discrepancy between it and the other face, constrained by the feature correspondences established between the images.

Perrett's model of sexual dimorphism includes images of one averaged male face and one averaged female face. Using warping, the appearance of the male average can be "feminized" by making its structure more like that of the female average. Interestingly, the appearance of the male average can be "masculinized" by making its structure *less* *like* that of the female average. Perrett's group justified the model's validity by referring to the changes in terms of verbally described features such as jaw size and lip thickness (Little, Penton-Voak, Burt, & Perrett, 2002).

In the face perception literature, a metaphor called "face-space" is used to explain face encoding and recognition. It is a high-dimensional geometric space in which the axes describe the sources of variation that differentiate faces (for example, configural features). Because of the correspondence between the perceptual space individuals use to represent faces and the actual physical features that differentiate faces, face space is also helpful to describe theories of attractiveness that contend with morphological variation in faces. In terms of the face-space metaphor, if the male and female averages are single points separated by some distance, one can envision a path in face space from the male average to the female average (see Figure 6 below). Between the male and female averages lie feminized male faces and masculinized female faces. Beyond the female average are feminized female faces and beyond the male average are masculinized male faces.

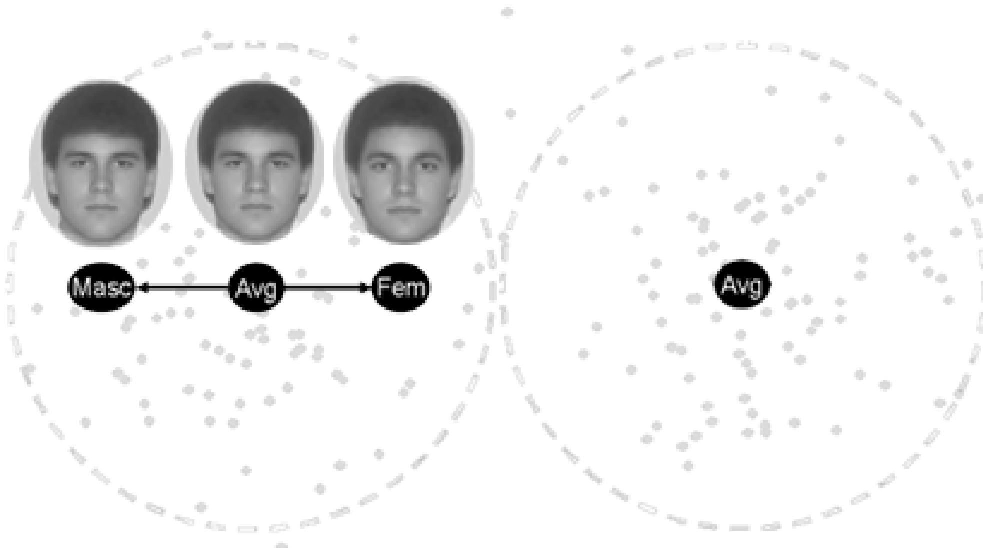

Figure 6: The male averaged face's hypothesized path through face space, when transformed by the female averaged face (right side "Avg").

Essentially, Perrett et al. (1998) created an image-based model of sexual dimorphism and within-sex variation in masculinity/femininity. They reasoned that the male and female average preserved hormonally-controlled sex-specific facial configural information. Therefore, creating new faces by *warping* the male and female faces varies the amount of sexual dimorphism, and thus the presumed amounts of sex hormones that would have been required to produce the faces if they were real (Perrett & Penton-Voak, 1999).

Research on the hormone marker hypothesis

Using stimuli created by warping male and female average face images, Perrett et al. (1998), found that men preferred feminized female faces but women preferred *feminized* male faces. The latter result was controversial based upon evolutionary psychological theory and other findings in the facial attractiveness literature. First, evolutionary psychological theories of attractiveness held that men's masculine features should be preferred to men's feminine features because of theorized associations between masculinity, testosterone, and immune system function (Gangestad & Thornhill, 1997; Thornhill & Møller, 1997). Second, most studies found consistently weak (r = .3) but *positive* associations between masculinity and attractiveness in men's faces (e.g., Brown, Cash, & Noles, 1986; Cunningham, Barbee, & Pike, 1990; O'Toole et al. 1998).

Perrett et al. (1998) realized that their findings were inconsistent with previous research and were, in fact, controversial - especially within the Evolutionary Psychological literature (e.g., Fink & Penton-Voak, 2002; Johnston et al., 2001). Little et al. (2002) found a way to minimize the controversy, by noting that there is a precedence for such difficult-to-explain findings, Cunningham's *Multiple Fitness Model*, in which attractive men's faces have a "combination of both masculine and feminine features, and so reflect 'multiple motives' in female mate choice" (2002, p. 67).

Perrett's research group also addressed the controversy by demonstrating that female preferences are contingent, depending on whether a woman is in a relationship, whether a woman desires a short-term relationship (Little, et al., 2002), and a woman's menstrual phase (Penton-Voak, et al., 1999). When women ovulate they prefer masculinized men, when they are less fertile they prefer feminized men. Perrett explained that women preferred the masculine men when they are ovulating so that they could be inseminated with the sperm of the men who have good genes, whereas women favor feminine men when they are less fertile, in order to capture the affections of men who are likely to take care of the children fathered by the masculinized men (Penton-Voak et al., 1999). This multiple-contingency approach helped integrate their findings with the rest of the Evolutionary Psychological literature (e.g., Buss sexual strategies – Buss & Schmitt, 1993; Cunningham's multiple motives – Cunningham, 1986) and it has inspired similar research (e.g., Gangestad, Thornhill, & Garver, 2002).

Johnston et al. (2001) proposed a somewhat different model of image morphing masculinization. They argued that Perrett et al.'s (1998) assumption that within-sex variation is an extrapolation of sex differences may not be correct and proposed a second morphing method claiming it more appropriately models the process of sexual differentiation. Johnston et al. proposed that, rather than caricaturing an averaged man with an averaged woman to produce a masculinized man, instead the end point of the morphing continuum should be artificially synthesized faces of a masculine man and a feminine woman. These synthesized images were produced by participants who used a program developed by Johnston to create an image of a man's face that appears very masculine. Thus, Johnston et al.'s model was represented as a movie of morphed transitions between the face images of a synthesized masculine man, to an averaged man, to an averaged woman, to a synthesized feminized woman; in terms of face space the model is three lines (between the four faces) joined at angles (see Figure 7 below).

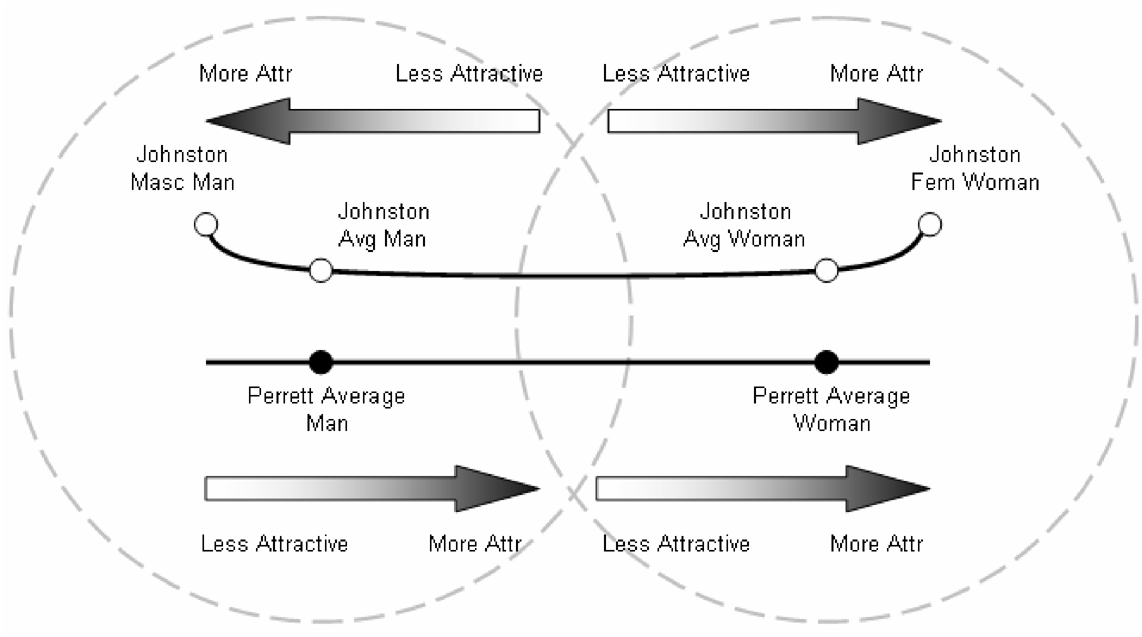

Figure 7: Schematic representation of the differences between Johnston's & Perrett's morphing models. The large dashed gray circles represent the boundaries of the positions of all men's (left) and all women's faces (right). The average faces are near the center of the clusters. The solid black lines represent the face space trajectories of the stimuli used by Johnston and Perrett. The arrows represent the relation of the stimuli to attractiveness by gender for Johnston (top) and Perrett (bottom).

Johnston et al. (2001) found that a sample of 42 women preferred masculinized men’s faces and feminized women’s faces, which partially conflicts with Perrett’s (1998) results. Figure 7 shows that Perrett’s results predict that feminized faces of both men and women are more attractive (arrows point away from men’s circle) whereas Johnston’s results predict that sexually dimorphic faces of both men and women are more attractive (arrows point away from circle centers). Interestingly, Johnston et al. (2001) found a preference for masculinized male faces and replicated the menstrual cycle dependency. Women in the ovulatory phase of their cycle more strongly preferred the masculinized men’s faces. These findings are in agreement with EP theory and most previous research.

Critique of the hormone marker studies

We could consider that the warped faces method eliminates noise implicit in other methods that use real faces (i.e., rating or measurement methods). The claims about the relationship between attractiveness and masculinity/femininity, however, theoretically extend to all faces; they are not limited to warped average faces. If warped face trajectories truly represent sexual dimorphism then they ought to be reducible to the dimensions differentiating real faces. Projected in face-space, Perrett's warping was a single trajectory between and beyond the male and female averages, whereas Johnston's was three trajectories joined at angles (see Figure 7). Left out of the projections into abstract and simple face space representations, however, is an understanding of where these hypothetical faces are in relation to the thousands of real faces that exist in the world. Speculatively, if femininity and attractiveness in women's faces are strongly correlated no matter how one determines the association (Cunningham, 1986; O'Toole et al., 1998; Rhodes, Hickford, & Jeffery, 2000), then perhaps it is reasonable to conceive of attractiveness as being collinear with a dimension that differentiates men's and women's faces. If, however, masculinity and attractiveness are weakly correlated (according to most methods they are), then maybe it is not justified to extend the relationship to men's faces; there may not be a single trajectory that "explains" attractiveness in terms of sexual differentiation for men. In fact, researchers have appeared to have identified at least two trajectories associated with attractiveness (Perrett et al., 1998; Johnston et al., 2001). Almost inexplicably, although the two trajectories have some visual equivalence, in that they both describe a masculine-feminine path through the space, they have opposite relations to attractiveness (see Figure 7).

Little et al.'s (2002) and Penton-Voak, et al.'s (1999) proposals that women's preferences are multiply contingent do not address the controversy mentioned earlier.

They found that women preferred feminized male morphed faces but most prior studies using *unaltered* faces found positive correlations between masculinity and attractiveness. Indeed, it is the generalization to unaltered male faces to which their research aspires. If women's preferences are multiply contingent the majority of women raters participating in previous studies would have had to have been ovulating, which is unlikely. Inter-rater reliability for ratings such as attractiveness and masculinity are usually very high when participants rate of a number of unaltered men's faces, even when men's and women's ratings are compared (e.g., Langlois & Roggman, 1990; Penton-Voak, et al., 2001), which makes doubtful the possibility that if researchers had noted participants' menstrual cycle phase and relationship status that they would have found reliable between-subject differences that support Perrett et al.'s contingency hypotheses. Furthermore, adding the behavioral contingencies requires that several elements and connections be added to the basic model proposed by evolutionary psychologists (see Figure 4) - facial features, sex hormones, and *personality traits* as well as situational connections to female preferences.

These extra degrees of freedom would complicate, rather than simplify, the subject. What Perrett et al. propose is that hormone action during puberty on growth that caused a person's face to be more masculine (or feminine) predicts their personality and behavior as an adult. A study to determine whether adolescent hormonal profile predicts adult personality has never been carried out because of its expense, complications, required time, and the implausibility of showing such effects of sex steroids on growth.

Rather than argue that Johnston's or Perrett's model more accurately represents sexual dimorphism Bronstad, Ramsey, and Langlois (2002) thought that a methodological artifact could explain the differences in the identified relationship between attractiveness and masculinity. They found strong support for the hypothesis that the number of faces used in a study accounted for differences in results. Studies that used fewer faces in identifying the relationship between masculinity and attractiveness (or femininity and attractiveness) were more likely to find discrepant results, which makes sense in light of the fact that confidence interval size is inversely proportional to sample size. Furthermore, almost all of the studies reporting a relationship between masculinity and attractiveness (including the divergent results of Perrett and Johnston) could be accounted for by a single hypothesis; masculinity and attractiveness are correlated (roughly) +.36. The number of faces used in each study was proportionate to the degree of disparity from the null hypothesis; the effect size of each study was within the 95% confidence interval of resampled correlations between attractiveness and masculinity made to a set of men's faces. It is therefore likely that methods that employ fewer stimuli will find discrepant results. In particular, studies using morphed faces have an unfortunate constraint that they use many variations of the same face image, and results of those studies are often discrepant.

A critique of evolutionary psychological assumptions of hormonal influences on growth

No analysis of the sexual dimorphism theories of facial attractiveness would be complete without discussing its most fundamental assumptions. Indeed, that a person's facial appearance reveals his or her hormonal history is fundamental to practically all EP facial attractiveness theories. For example, men appear more masculine because they have had more male hormones than men who look less masculine. Women prefer masculine men because the facial evidence that they have had more male hormones proves their genes are of good quality. The fundamental points of the hormonal hypothesis are: 1) Although men and women have the same types of facial features, many features have different male and female forms; 2) Starting in puberty, male hormones masculinize men's faces resulting in masculine features; female hormones feminize women's faces resulting in female features; 3) Male hormones, such as testosterone, promote growth, whereas female hormones, such as estrogen, inhibit growth: 4) Within each sex there is a monotonic relationship between the amount of sex-typed hormone and how masculine or feminine the facial appearance is. For example, men have longer chins than women. EP attractiveness researchers assume that men with longer chins have had more pubertal testosterone and less pubertal estrogen than men with short chins (Johnston et al., 2001; Perrett & Penton-Voak, 1999; Thornhill & Møller, 1997)

EP attractiveness researchers are not often explicit about the relationship between growth and hormones, usually stating that hormones "influence" or "affect" the facial features. The theory, however, is more straightforward and complements research methods best if a monotonic or linear relationship is assumed. For example, Johnston et al (2001) used a sequence of images, a movie in which an average male face changes into a very masculine male face, to represent the differences in hormones that would have been required to produce the faces if they were real. They claimed that:

> "…because such secondary sexual characteristics are mainly a consequence of different levels of pubertal hormones, this methodology provides a basis for interpreting how facial preferences are related to the degree to which such hormonal markers are displayed on the faces of men and women." (Johnston et al. 2001, pp. 261-262)

Johnston et al's claims indicate that they assume the relationships among facial features, hormones, and attractiveness preferences to be highly related. Moreover, Perrett and Penton-Voak stated that the difference between men's and women's faces "parallel the differences between individuals with high and low androgen levels" (Perrett & Penton-Voak, 1999, p. 662).

Unfortunately, the monotonic hormone theory is unsupported by evidence. First, although estrogen and testosterone may be some of the factors responsible for between-sex differentiation (Grumbach, 2000), they are not responsible for within-sex variation.

Second, *between*-sex differentiation can not be attributed to estrogen and testosterone in the simple manner assumed (i.e., testosterone facilitates growth, estrogen inhibits growth). Grumbach states that the “belief that the human male skeleton accrues greater bone mass than that of the female because of the action of testosterone no longer appears acceptable” (p. 258). Grumbach points to the increasing evidence showing that estrogen has a critical role in masculinization of males.

Causality in growth is very complicated. Part of the complexity of hormonal influences on growth is evident in the writings of the foremost authorities on growth. For example, Enlow, an authority on cranial growth, wrote that by the end of the 1980s, growth researchers realized that the established theories of growth, which tended to be “straightforward, [and] easy to understand,” were unsound (1990, p. 229). Tanner (1990) pointed out that a hormone may increase the sensitivity of a second hormone’s receptor, an event difficult to discriminate from an increase in the second hormone alone. Hormones also self-prime, causing an “explosive interaction” (p. 85) that eventually results in a loss of sensitivity (i.e., down-regulation). In Tanner’s writings, there is scant information about how hormones cause within-sex variation, primarily because the data do not suggest any relation. For example, with regard to growth hormone (GH), studies do not show that tall children secrete more GH than average or short children. According to Tanner, "the endocrine problems of growth seem not, after all, to be entirely solved” (Tanner, 1990, pp. 91).

The idea that estrogen inhibits bone growth is important to EP face attractiveness theory. Consistent with this hypothesis, Tanner writes that estrogen facilitates epiphyseal fusion. *Epiphyses* are active areas of growth until “capped” by estrogens. Estrogens, however, facilitate growth as well; estrogen inhibition on growth is due to a self-priming process, rather than being its only effect. Recent evidence suggests estrogen, rather than testosterone, initiates adolescent boys' pubertal growth spurt (Grumbach & Auchus, 1999). Thus, estrogen and testosterone at time 1 can have a permissive effect, and at time 2 have an inhibitory effect.

The insistence on explaining variability in sex-typical facial variation by hormones also ignores genetic causes of variability not strictly due to testosterone or estrogen. There are at least three types of genetic causation: 1) Mendelian, such as eyebrow peakedness (Itin, Kirtschig, Gilli, & Happle, 1997) and variation in tongue-rolling (McKusick, 1992, the authoritative human Mendelian traits catalog); 2) Early development, for example, hox genes control basic aspects of body plan (Thesleff, 1997), and; 3) Polygenic causation, for example, Hunter, Balbach, and Lamphiear (1970) found, in a sample of 38 families, that mandibular length of fathers and sons, as well as fathers and daughters, was correlated +.6. Height of face from eye to chin was correlated in fathers and daughters (.46), but not fathers and sons (.32, *ns*). Saunders, Popovich, and Thompson (1980) found only small differences between maternal and paternal influence on masculine traits, such as mandibular length and lower facial height, in 147 families.

In summary, we can not understand the historical hormonal profile of individuals, compared to others of the same sex, based upon analysis of their faces. It may be that there are "natural experiments" that afford us an educated guess in special circumstances (e.g., men with Kleinfelter's disorder), but EP face attractiveness theories don't deal with special circumstances; their scope is human psychology and everyday variation.

EP borrowed many ideas and methods from Ethology. One of the most influential Ethologists, Tinbergen, however, insisted that our theories be influenced by our knowledge of biological mechanism and development (e.g., Tinbergen, 1976). If so, then we should consider developing a new theoretical structure that is consistent with what we know of physiology and face perception. It is possible to construct theories of facial attractiveness that involve sexually dimorphic features without requiring an immuno-endocrinological foundation. For example, Enquist et al. (2002) hypothesized that preference for masculine men's or feminine women's faces are due to a cognitive strategy to discriminate men's and women's faces.

The contribution of these features can be assessed by focusing on the aspects of faces germane to the particular theories. For example, if humans use sexual dimorphisms to make judgments of attractiveness, hypothesis tests can involve representing variance in sexual dimorphism and determining how it affects judgments of attractiveness, rather than measuring apparently adaptive behaviors and claiming consistency with hypotheses involving unobserved hormonal causes. The purpose of my first study is to test sexual dimorphism theories of attractiveness. I will describe a new method of transforming face images that will be used to test how masculinization and averageness (i.e., prototypicality) affect facial attractiveness.

AVERAGENESS

The averageness hypothesis is one of the few facial attractiveness hypotheses not typically formulated in terms of immuno-endocrinological good genes theories (Langlois & Roggman, 1990). Not accidentally, it is the only hypothesis of facial attractiveness that posits testable cognitive claims rather than simply explaining variance in attractiveness preferences with variation in facial appearance. For example, Rubenstein, Kalakanis, & Langlois (1999) found that infants react to facial composites as if they have seen them before if they have previously viewed the faces from which the composites were generated.

An advantage of the averageness hypothesis is it predicts that the entire face is important to facial attractiveness, meaning that averageness is potentially fundamental and necessary for facial attractiveness (Rubenstein, Langlois, & Roggman, 2002, p. 21).

Individual faces are less “facelike,” or less like a prototypical or average face, are predicted to be less attractive. According to the hypothesis, individuals abstract the central tendency of the category (and subcategories) of faces, represented neurally as a face prototype. The similarity of a face to a facial prototype determines its attractiveness; faces more similar to a prototype are more attractive than less prototypical faces. Fundamentally, the averageness hypothesis depends on knowledge of faces that develops through experience, and the ability to compare face representations (Rubenstein, Langlois, & Roggman, 2002; Rubenstein, Kalakanis, & Langlois, 1999). The face perception system relies on these capacities for detection, classification, and recognition.

Averageness: Experimental studies

The first psychological test of averageness theory was Galton’s observation that when images of criminals are combined via timed-exposure photography, the resulting photograph doesn’t reflect a distillate of criminality, but it happens to be handsome (Galton, 1879). Langlois and her colleagues (Langlois & Roggman, 1990; Langlois, Roggman, & Musselman, 1994) explored the averageness effect in more detail, using image digitization and manipulation technology that allows for better production of averages, or face prototypes, eliminating the multiple outlines observed in multiple-exposure composite photography. They found that they could make face averages more attractive by using more faces to construct the average. That is, participants found composites created from 16 faces more attractive than composites created from 8 faces, which were more attractive than composites created from 4 faces. There was no significant increase in attractiveness when the number of faces combined was greater than 32. This finding demonstrates the feasibility of a cognitive explanation of attractiveness, based on the prototype theory of categorization. If attractiveness judgments depend on a comparison of a face to an abstracted face prototype, then a simulated prototype face should itself appear attractive.

Critics of Langlois and Roggman countered that averaged faces were attractive because they are symmetric and/or blurred (for example, Alley & Cunningham, 1991). Langlois, Roggman, and Musselman (1994) answered these criticisms and others, showing that "soft-focus photographic" effects don't account for increased attractiveness of averaged faces; if different images of the same individual are averaged together the resulting averaged image is blurry but not more attractive. Additionally, Langlois, Roggman, and Musselman described experimental and correlational evidence that symmetry and attractiveness are unrelated.

Related to averageness, some research in the face perception literature centers on the construct called typicality. Typicality of faces is related to both how well faces are remembered and their averageness. The typicality of a face also predicts how quickly people can classify it as a face (Valentine & Bruce, 1986b). Faces rated as more typical are easier to classify as faces but are more difficult to recognize as a previously seen individual (Valentine & Bruce, 1986a). The favored explanation is that the distribution of faces within the stimulus space accounts for typicality effects; typical faces are closer to a central cluster in which crowding effects make recognition difficult but ease category classification (Johnston & Ellis, 1995).

Researchers explain and conceptualize typicality and other facial cognition phenomena with the face space metaphor. Face space is a high-dimensional geometric space, the axes of which are the factors that differentiate faces, faces are single points, and the distance between faces corresponds to similarity – similar faces are close to each other. Any measure of similarity can be placed into the framework by submitting similarity scores to principal component analysis or multidimensional scaling (MDS).

Moreover, researchers informally conceptualize the relations among faces in abstract face spaces, for example, speculating that men's and women's faces form separate clusters within a common space (see Figure 8). Typically, (hypothetical) averaged faces are at the center of the clusters (Valentine & Bruce, 1986a; Busey, 2001).

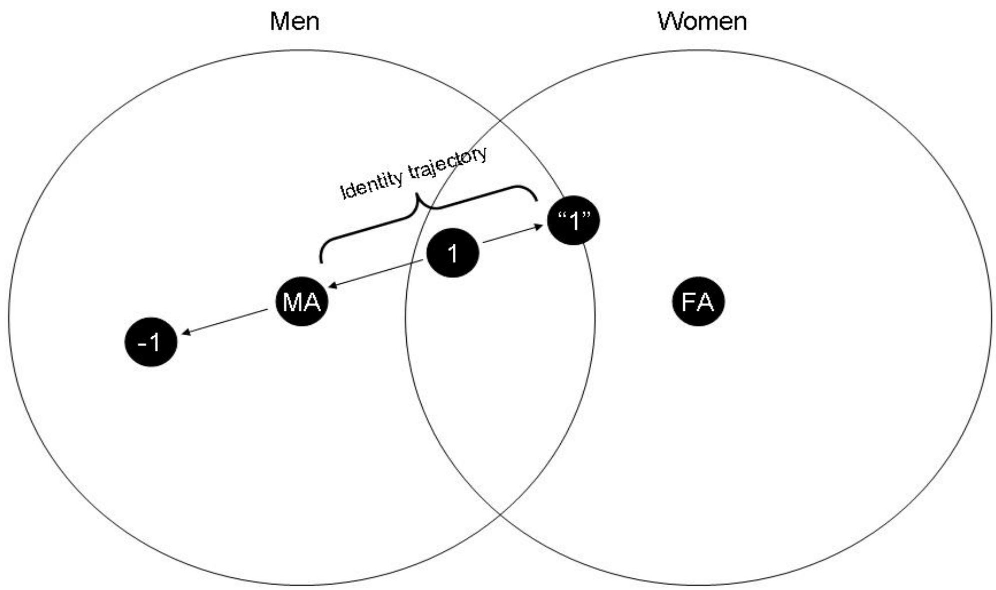

Figure 8: 2-dimensional abstract face space. The large unfilled circles define boundaries for locations of men's and women's faces. MA and FA show the theoretical location of male and female averages. The circle marked 1 is an individual face of ambiguous gender. "1" is its caricature, -1 is its anti-face (described below).

Face image morphing is an image manipulation technique that has been described as "navigating face space" (O'Toole, Wenger, & Townsend, 2001). The person morphing faces must establish a correspondence between two faces' features by placing points on similar locations of each face, such as around the eyes, nose, mouth, outline; (Figure 8 signifies that faces 1 and "MA" are placed in correspondence). When the set of feature correspondences between two faces is established, one face may be manipulated by exaggerating or minimizing its differences with the other face. In terms of Figure 8, using the male average face (MA) as a reference, face 1 can be manipulated to appear less like or more like MA.

When a face is morphed with an average face, its structural changes cause interesting corresponding psychological effects. As it becomes more like the average face, it is rated as more attractive, and more typical, than the original. The altered face is also recognized as the original face less often as it becomes more average (Lee, Byatt, & Rhodes, 2000; O'Toole, Price, Vetter, Bartlett, & Blanz, 1998; Rhodes, Sumich, & Byatt, 1999). As it becomes less like the average face it is rated as less attractive, less typical, and it is identified as a "better example" of the face than the original. By analogy it is suggested that "identity" of individual faces radiates from the prototype like spokes from the hub of a wheel. Thus, computational models based on the prototype hypothesis of face categorization may use the angle, in face space, between face "spokes" as the measure of similarity between faces (Busey, 2001).

Recently discovered sensory adaptation illusions provide interesting evidence for the role of a prototype in facial cognition (Leopold, O'Toole, Vetter, & Blanz, 2001; MacLin & Webster, 2001). MacLin & Webster found that participants who adapted to distorted facial images perceived undistorted face images as distorted in the opposite direction. That is, if they presented images of faces that they stretched vertically, after adaptation the distorted faces appeared normal whereas undistorted face images appeared distorted, as if they were compressed vertically. Leopold, O'Toole, Vetter, & Blanz found that such adaptation effects extended to faces that were not grossly distorted, and that sensory adaptation could be used to systematically alter perception of facial identity. The trajectory between a face and the average face contains information distinguishing that face from other faces (i.e., its identity). If the face is morphed with the average so that it looks more like the average face, it begins to be less recognizable. *Anti-faces* exist beyond the trajectory between the individual face and the average, projecting from the average face and distant from the individual face. Anti-faces differ from the original face in many ways that faces can differ from each other. Leopold et al. used these anti-faces in an experiment in which participants learned to associate names with several face images. For each veridical face image the experimenters constructed a contrasting anti-face. Participants adapted to an anti-face for 5 seconds, and then viewed a different face for .5 seconds. If presented the average face after adapting to an anti-face, they would mistake the average face for the anti-face's complement. In other words, after adapting to "Anti-Bob" they then briefly viewed the average face, who they mistook for Bob.

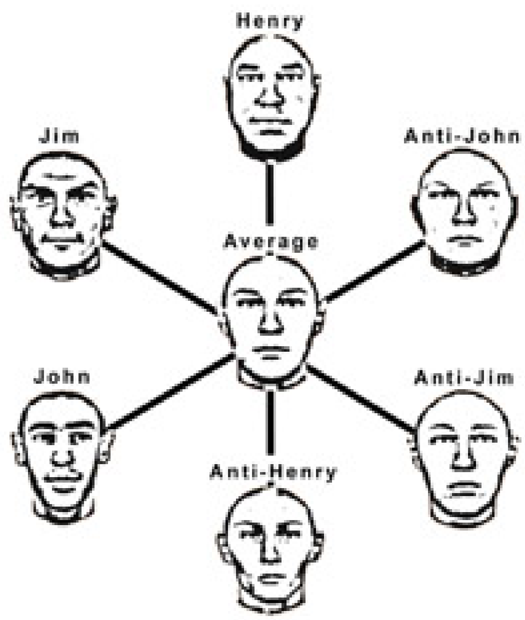

Figure 9: Faces and their anti-faces (after Leopold, O'Toole, Vetter, & Blanz, 2001).

These short-term sensory adaptation effects suggest the presence of cortical receptive fields selective for configural variation in faces; neuron populations adapt to viewed faces, which decreases activation for similarly structured faces and increases activation of receptive fields selective for oppositely structured faces. Moreover, this study suggests that a “prototype” is the consequence of the distribution of cortical receptive fields; such adaptation effects would not be found if cortical face representations were not organized somewhat like theoretical face spaces that have a prototype at the centroid.

Averageness: Correlational studies

If averageness defines attractiveness then it is important to test whether attractive faces are more similar to average faces than are unattractive faces. This hypothesis test has been conceived of as determining whether facial feature distances measured from attractive faces are closer to mean values than those of unattractive faces, a method which I have already criticized in this dissertation. For example, Grammer & Thornhill (1994) measured several feature distances of 16 men’s and 16 women’s face images. They found that measured averageness did not predict men’s attractiveness and negatively predicted women’s averageness. They also noted that the averaged faces they measured did not appear to have more average facial proportions by their measurement method: “computer averaging does not necessarily create higher metrical averageness in faces…” (p. 237). This result is awkward for Grammer & Thornhill; if averaged faces can not be distinguished from unaltered faces, the measurement technique is not useful to measure averageness.

Jones (1996) used a facialmetric approach in which he evaluated the averageness of 16 facial distance measures combined in two different ways. His study was cross-cultural, using face photos and raters from five different cultures (Ache and Hiwi Native American tribes, Brazil, United States, and Russia). Jones found only weak support for the averageness hypothesis due to significant, but small, correlations between metric averageness and attractiveness for the face images of the Ache. In none of the other 4 populations was the relationship significant. The results of the study could be problematic for the averageness hypothesis.

In the third study in which averageness was measured as a combination of linear feature distances, Pollard, Shepherd, & Shepherd (1999) measured 19 feature distances from pictures of 10 men's and 10 women's faces. They defined facial averageness as the summed z-scores for all distance measurements. Pollard et al. asked participants to rate the attractiveness of the images and found that facial averageness was unrelated to attractiveness. Pollard et al. concluded that faces average in proportion are not unusually attractive.

The three facialmetric tests of the averageness hypothesis all have the same methodological problem. As mentioned already, face distance measures are insufficient for modeling facial representation. When a similar number of distance measures are combined according to a model that weights each feature optimally - rather than combining features as if each is interpreted exactly like the other – human like performance can't be achieved (Burton, Bruce, & Dench, 1993). Furthermore, face recognition ability is impaired when individuals to be recognized are represented by line drawings (Davies, Ellis & Shepherd, 1978). Therefore the methods of Grammer and Thornhill (1994), Jones (1996), and Pollard, Shepherd, and Shepherd (1999) are not appropriate tests of averageness theory.

Critique of averageness studies

Whether attractive faces are average is crucial to testing the averageness hypothesis, but the investigation should be framed carefully. The question of whether attractive faces are metrically average involves two types of sampling problems already discussed in relation to other theories of facial attractiveness. The first sampling problem is which faces researchers use to test the hypotheses. The second problem is which facial features researchers use to test the hypotheses. There are so many possible features to measure (Farkas, 1981), researchers could consider feature selection as a random sampling process, just as establishing a connection between averageness and attractiveness involves selecting a sample of faces.

The alternative to facial feature sampling is to use a model of face perception to measure faces. Perception and measurement are often equivalent processes; each is a type of stimulus representation (Edelman, 1998). The practice of perceptual modeling is essentially that of building a reflection of the world as embodied by the perceiver. Thus, the act of modeling the perceiver's world also produces a model of the world (Palmer, 1975). Ultimately, facial attractiveness researchers wish to explain the experience of the perceiver, so in answering the question of what makes faces attractive, the choice of method – sampling or modeling – is straightforward: we should model the perceptions of our participants.

O'Toole, et al. (1998) authored the only study in which a cognitive model of face perception was used to predict attractiveness ratings of faces. They used principal component analysis (PCA) to quantify averageness. PCA has been incorporated into automated face recognition tools that take raw images as input and compare them to previously stored faces (Turk & Pentland, 1991). It shares some qualities with face perception theories, such as a multidimensional feature space, holistic receptive fields, and a central prototype (Cottrell, Dailey, Padgett, & Adolphs, 2001; O'Toole, Wenger, & Townsend, 2001). O'Toole et al.'s test of the averageness hypothesis was essentially how strongly a face activated the receptive field of the prototype, which is not an unreasonable test of the averageness hypothesis. More specifically, O'Toole, et al. used each face image's loading on the first eigenvector of a principal components analysis as a measure of averageness. This variable is each face's loading on the largest axis of pixel variation in the entire set of images and, in this particular context, corresponds to the average face of the stimulus set. They found that this variable weakly predicted men’s, but not women’s, attractiveness. This method is arguably the most interesting measurement method used in the literature; it addresses configural features and is connected to the face perception literature. It is a radical step forward, and can be extended.

The purpose of my third study is to use a face perception model as a measurement method and test of averageness theory, replicating and extending the findings of O’Toole et al (1998).

Chapter 3: *Tests of the Theories*

There are two tenable theories of facial attractiveness: sexual dimorphism and averageness. The experiments forming this dissertation were designed to provide critical tests of each theory. The first experiment contrasted averageness theory with sexual dimorphism. The second study was inspired by the results of Study 1, a computational model designed to determine whether faces manipulated to be more masculine in appearance differ importantly from faces that vary naturally in masculinity. The third study tests averageness with two computational implementations of the hypothesis.

EXPERIMENT 1: TESTS OF AVERAGENESS AND FEMINIZATION HYPOTHESES

Perrett et al. (1998) observed that when they feminized an averaged male face by warping it to look more like an averaged female face it was rated as more attractive than masculinized male faces. They claimed that this finding showed that non-average faces were more attractive than averaged faces and, therefore, that their research "refutes the averageness hypothesis" (p. 885). Perrett et al., however, used only slightly altered versions of an averaged face and all of the stimuli were highly similar in facial appearance (see Figure 3). Thus, the refutation is premature because the stimuli look to be different photographs of the same face rather than different faces.

It is likely that averageness and sexual dimorphism are both fundamental dimensions of the stimulus space that describes facial variation. Whether facial averageness or sexual dimorphism explains attractiveness, however, has not been addressed by an adequate experimental design. This experiment seeks to remedy that by providing a context in which the relative contribution of averageness and feminization transforms can be observed.

As mentioned earlier, researchers alter the averageness of a face by warping or morphing a picture of an individual face (face 1 in Figure 8) with the image of an averaged face (Rhodes, Sumich, & Byatt, 1999; O'Toole, Price, Vetter, Bartlett, & Blanz, 1998). One can alter face "1" to be more like the averaged face, which tends to generate higher attractiveness ratings than the original, or to be less like the averaged face (face "1" in the Figure), which tends to generate lower attractiveness ratings than the original. Transforming a face to look less like the averaged face is called caricaturing because it augments the idiosyncrasies within the transformed face, which is approximately how caricatures such as those in political cartoons are envisaged (Gibson, 1969, p. 102). Anti-caricaturing is the reverse operation; it deemphasizes the individuating information by making the face image look more like the averaged face.

Researchers alter the masculinity or femininity of a face image by first creating an averaged man's and an averaged woman's face image. To "masculinize" a facial image, one uses image warping to make it to look less like the female average, which removes female facial characteristics and amplifies idiosyncrasies present in the warped face (caricaturing). To "feminize" the image of a face, one warps it to look more like the female average (anticaricaturing), or incorporating feminine facial characteristics. Both men's and women's faces, when feminized, have been proposed to be more attractive than the original faces (Perrett, et al., 1998).

Both averageness and feminization theories can explain what it is about any face we see that makes it attractive or unattractive. A major difference in how the theories have been tested, however, is the variety of faces over which the effect has been observed. We have criticized the feminization research on these grounds (Bronstad, Ramsey, & Langlois, 2002); researchers identifying a negative relationship between attractiveness and masculinity have used uncommonly small samples of contrived faces.

Several studies suggest that the masculinity and attractiveness are only weakly correlated, so different, small, samples of men's faces could produce inconsistent results. When researchers estimate the correlation between attractiveness and masculinity ratings by using a moderate-sized random sample of real face images, they find that the association of attractiveness and masculinity of men's faces is positive, contrary to the feminization hypothesis (Brown, Cash, & Noles, 1986; Cunningham, Barbee, & Pike, 1990; O'Toole, et al., 1998; Penton-Voak, et al., 2001; Scheib, Gangestad, & Thornhill, 1999). Interestingly, no comparable controversy occurs over women's faces; the correlation between femininity and attractiveness ratings of women's faces is so strongly positive that it is virtually impossible to draw a sample in which the association is negative. Thus, it is likely that either small sample size or the practice of warping faces has confused the relationship of masculinity and attractiveness in men's faces.

One way to resolve the problem of sample size is to obtain many male faces and transform them into versions that are more masculine or more feminine using an extension of the technique that Perrett et al. (1998) developed. There are two problems with feminizing or masculinizing ordinary faces. First, averaged faces are not extremely masculine or feminine. Referring back to Figure 8, morphing face "-1" to look more like the male average will make it appear less masculine, because this will move it towards the female cluster. Second, it can appear more masculine by making it less like the female average face, but then it should become less average as well (it is being anticaricatured). In this scheme masculinization is confounded with averageness.

The way to resolve both problems is to use masculinized male and feminized female averages as tools to transform individual faces. The benefit of this method is that we can transform many faces to be more or less feminine and more or less average, and the modifications are essentially orthogonal to each other. Moreover, as the transformations are orthogonal this method is a simultaneous test of the averageness and feminization hypotheses. Figure 10 shows a conceptualization of the transformations.

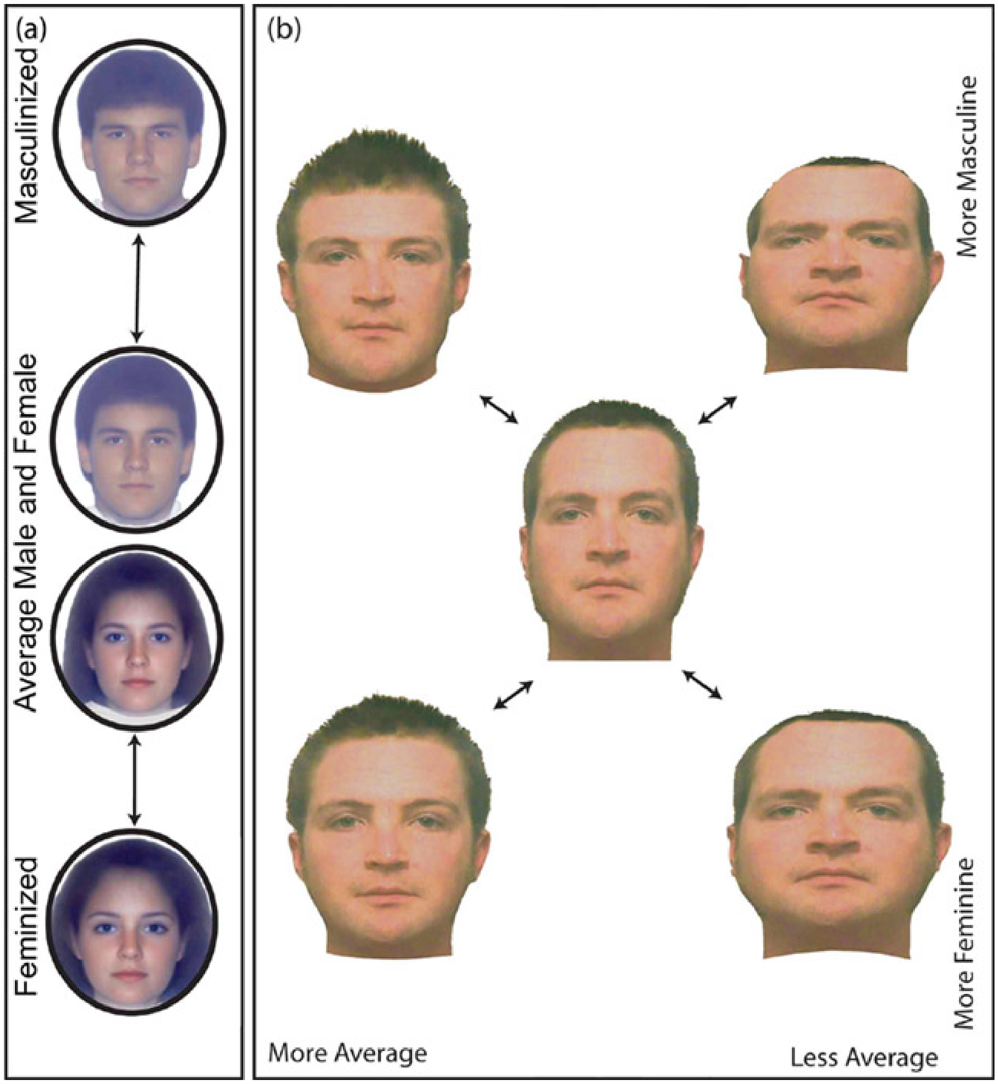

Figure 10: Schematic of image warping in Experiment 1.

In Figure 10, the original face is at the center of panel b. Four versions of the original were produced by warping it with the masculinized male average (top left of panel a) and the feminized female average (bottom left of panel a). The versions shown here are exaggerated for illustration – images used in the experiment are less strongly warped. To ensure that masculinization and feminization made faces more masculine and feminine, respectively, some participants rated the images for masculinity.

Method

Stimuli

Images of 8 men's and 8 women's faces were transformed by a computer warping algorithm (Gryphon Morph™). Faces were transformed to be: 1) less average and more feminine; 2) more average and more feminine; 3) less average and more masculine, and; 4) more average and more masculine. There were two sets of stimuli rated by two groups of participants. The first group of participants rated 64 stimuli (images of 16 individuals changed to produce 4 versions of each). The second group of participants rated subsets of 128 stimuli (images of 16 individuals given the 4 types of image warping at two different levels or strengths).

Gryphon Morph™ was used to perform the image warping. It allows users to change images to varying degrees, from 0% (no change) to 100% (the image of the individual would be changed to conform to the target image completely). Warping was performed at 40%, towards or away from the feminized female average and masculinized male averaged face images, for the first group of participants who viewed 64 stimuli. Additional versions of the original face images were created by warping the images at 20% to create all 128 images. Figure 10 shows a much stronger amount of warping of a single face for illustration.

Participants

81 people (49 men and 32 women) of college age participated, viewing the 64 faces transformed 40%. 107 people (73 men and 34 women) of college age participated in a follow-up study in which they viewed subsets of all 128 face images (across

participants, all images were rated). Participation was in partial fulfillment of an introductory psychology class requirement.

Procedure

Within the first group of participants, half (selected at random) rated faces for attractiveness on a 100 point Likert scale (very unattractive to very attractive). A second half of the first group of participants rated men's faces for masculinity or women's faces for femininity on 100-point Likert scales (very masculine to very feminine). Participants viewed images of faces on a computer screen, using a computer mouse to manipulate a graphic user interface slider. For both the attractiveness and masculinity/femininity ratings, the slider was programmed to relocate itself to a random position on the 1-100 scale after each rating was made. A block design was employed in which participants viewed and rated each of the 4 transformed versions of each face serially before they were presented with altered versions of a different face. Presentation order was randomized between- and within-blocks. Ratings were analyzed for interrater reliability (Cronbach's alpha).

Predictions and analyses

It isn’t clear whether the hormone marker hypothesis predicts that feminization or masculinization will be associated with increases in attractiveness for men’s face images. Usually researchers identify positive correlations between rated masculinity and attractiveness in men’s faces (e.g., O’Toole et al. 1998). In studies that used image warping, however, feminized men’s faces are more usually more attractive than masculinized men’s faces (Perrett et al., 1998; Rhodes, Hickford, & Jeffery, 2000; the exception is Johnston et al., 2001). The averageness hypothesis makes a clear prediction: making faces more similar to the average face will make them more attractive.

Accordingly, we should observe significant main effects of averageness and either masculinization or feminization on attractiveness.

A generalized linear model was used to determine the relative contribution of averageness and feminization to attractiveness ratings in men's and women's faces. Participant's masculinity/femininity ratings were examined to determine whether masculinization and feminization affected perceived masculinity and femininity in the expected manner. Specifically, if increasing the femininity of faces increases their attractiveness, then rated femininity should also be positively correlated with attractiveness.

Results

Participants preferred images of both men's and women's faces that were transformed to be more average, $F(1, 2800) = 237.10$, $p < .001$, compared with images transformed to be less average. Face images that were feminized were more attractive than those that were masculinized, $F(1, 2800) = 23.67$, $p < .001$. Thus, both the averageness and sexual dimorphism hypotheses were simultaneously supported. Participant gender had a small but significant effect on attractiveness ratings; women tended to give higher attractiveness ratings than men, $F(1, 2800) = 33.13$, $p < .001$. Figure 11 shows the mean attractiveness ratings participants assigned to the manipulated face images. The error bars indicate that perceived attractiveness varied across different faces within each transformation condition. That is, the error bars and means stand for attractiveness ratings given to each of the 8 men's face images within a transformation condition. The size of the error bars indicate that some of the men were rated as very unattractive and some were rated as very attractive, not that the effect of feminization was nonsignificant.

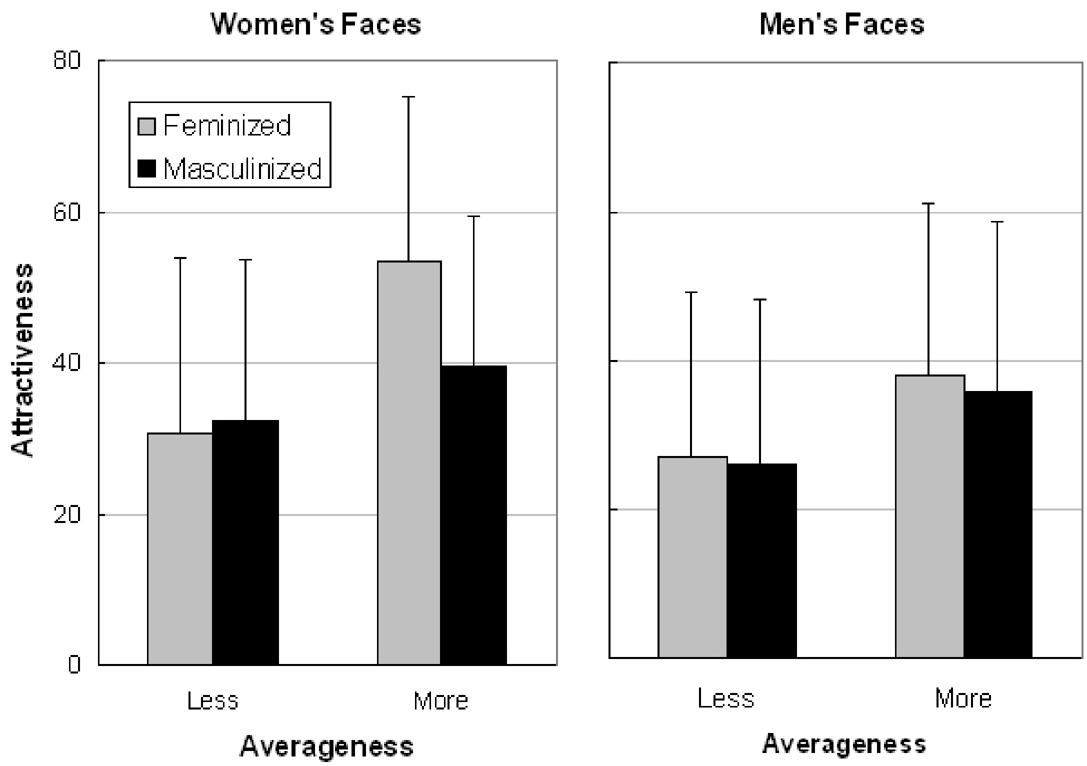

Figure 11: Attractiveness ratings given to transformed men's and women's face images.

For men's face images, masculinizing significantly increased ratings of masculinity, $F(1, 1124) = 81.70$, $p < .001$, whereas changing the averageness of faces did not, $F(1, 1124) = 2.06$, *ns*. For women's faces, feminizing increased femininity ratings, $F(1, 1148) = 94.43$, $p < .001$, and so did averageness transformations, $F(1, 1148) = 23.41$, $p < .001$. The effects in women's faces were qualified by a significant interaction of averageness and feminization, $F(1, 1148) = 9.57$, $p = .002$. Averageness transformations increased perceived femininity more strongly when women's faces were feminized than when they were masculinized.

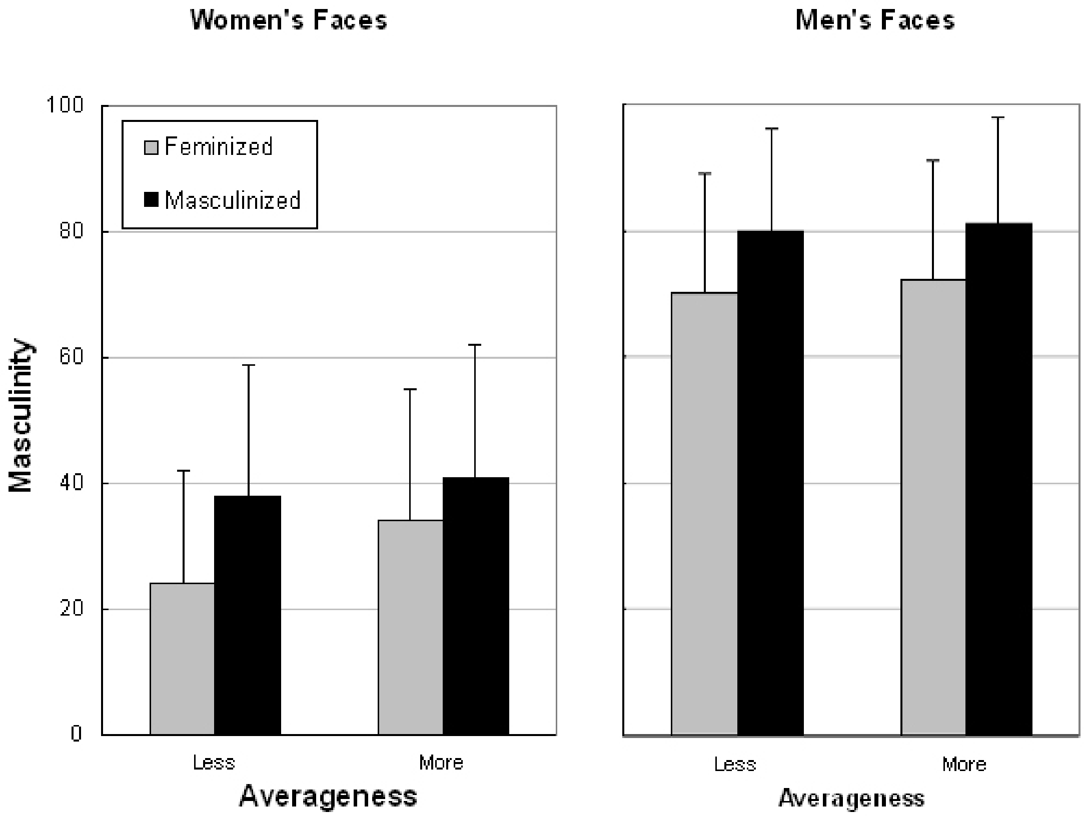

Figure 12: Masculinity ratings given to transformed men's and women's face images.

As participants made ratings of masculinity to the faces, it was also possible to examine the relationship of perceived masculinity to attractiveness across the transformed faces. For the women's face images, perceived femininity and attractiveness were positively correlated, r = .79, which is congruent with the finding that feminization increases attractiveness as well as previous findings that perceived femininity and attractiveness are strongly positively correlated (O'Toole et al., 1998; Rhodes, Hickford, & Jeffery, 2000; Perrett et al., 1998; Johnston et al., 2001; Brown, Cash, & Noles 1986; McArthur & Berry, 1987).

The picture is rather different for men's faces. Because femininization caused men's faces to be perceived as significantly more attractive, it would seem that perceived masculinity and attractiveness should be negatively correlated across the transformed men's faces. Instead, the relation was positive in both masculinized men's, $r = .46$, $p = .06$, and feminized men's faces, $r = .20$, $p = .46$. As these correlations were calculated across a relatively small number of face images, I examined the ratings participants gave (on 5-point Likert scales) to a larger number of transformed men's faces for perceived masculinity and attractiveness, the four transformation types were performed at 20% and 40% to double the number of images. The correlation between the two variables remained positive for both masculinized, $r = .66$ and feminized men's faces, $r = .62$ ($p <.001$ for both). One outlier was removed, with the outlier the correlation between masculinity and attractiveness for masculinized faces did not change much $r = .66$, $p <.001$. The slope of the regression line between masculinity and attractiveness for masculinized faces was steeper without the outlier ($\beta = .95$ without the outlier, $\beta = .71$ with the outlier included. The slope of the regression between masculinity and attractiveness for feminized images of men's faces was .85. The slopes did not differ significantly with or without the outlier included (Figure 13 shows slopes without outlier); the confidence interval for $\beta_{feminization}$ was .65 to 1.05, which includes the slope estimates for masculinization with and without the outlier.

The apparent paradox, that both feminization and masculinity cause men's faces to be evaluated as more attractive, is readily explained. Eight different men's faces were feminized, masculinized, and so forth, but within each transformation type the men whose faces were originally more masculine were always more attractive than those men whose faces were originally more feminine. These results indicate that feminization does not substantially alter the relationship between perceived masculinity and attractiveness; rather, it raises the regression line (see Figure 13).

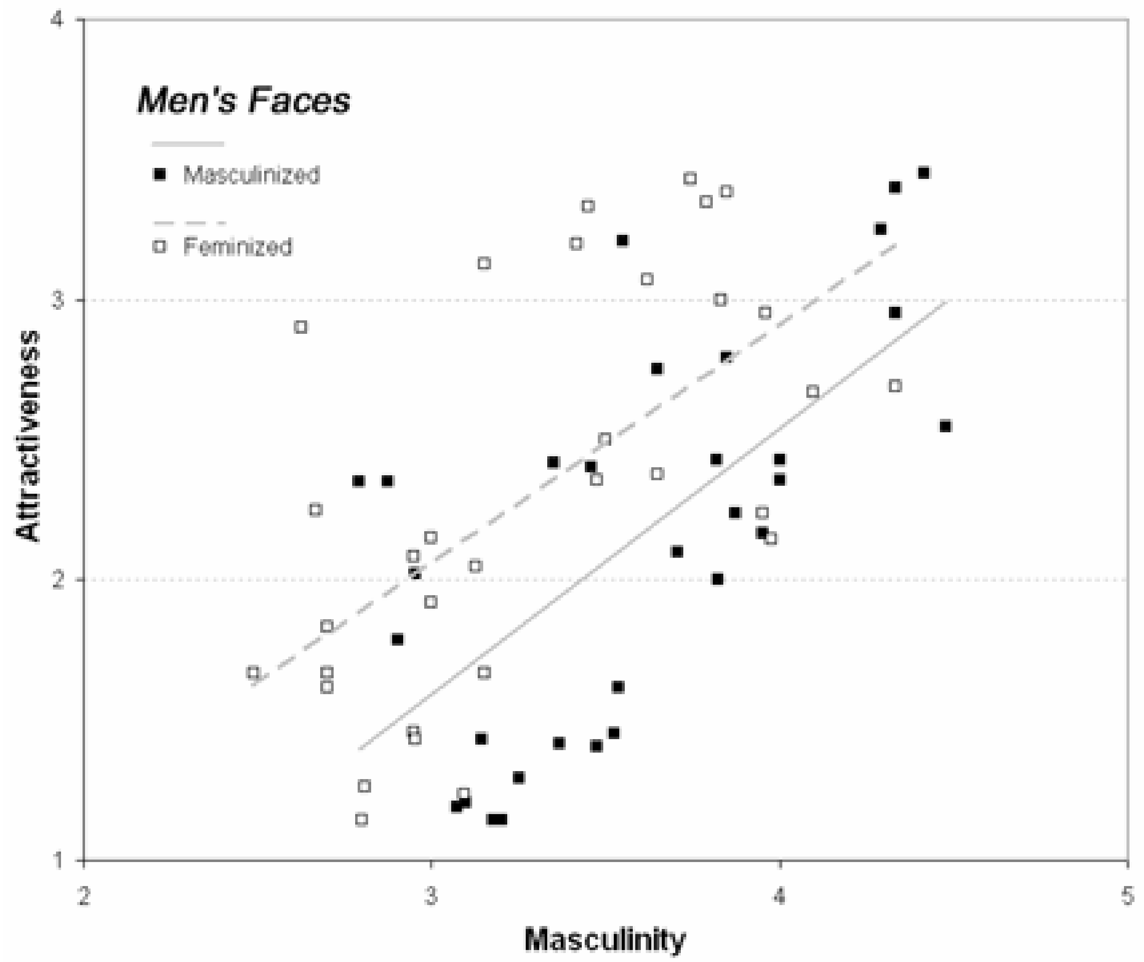

Figure 13: Perceived attractiveness and masculinity ratings of feminized and masculinized men's face images.

Discussion

There are two clear findings. First, averageness and sexual dimorphism appear to be independent explanations for facial attractiveness. Second, artificial feminization is positively, but perceived femininity is negatively, associated with judged attractiveness of men's faces. The paradoxical result that both masculinity and feminization increase men's facial attractiveness, explains the confusing discrepancies in the facial attractiveness literature; some researchers found that feminized men's faces are more attractive than masculinized men's faces (Perrett et al., 1998; Rhodes, Hickford, & Jeffery, 2000), whereas other researchers who analyzed perceived masculinity (Brown, Cash, & Noles, 1986; Cunningham, Barbee, & Pike, 1990; O'Toole, et al., 1998), or who tried to measure facial masculinity (Penton-Voak et al., 2001, Scheib, Gangestad, & Thornhill, 1999), found masculinity and attractiveness to be positively related.

The relation between facial structure and perceived masculinity is more complicated than it initially seemed. Penton-Voak and Perrett stated that their morphed faces "...embody the 'psychological' meaning of masculinity..." (1999, p. 662). Whereas masculinity and masculinization were thought of as interchangeable, these results show that they have different perceptual effects, at least concerning men's attractiveness. Perrett et al. (1998) also assumed that the changes in facial appearance induced by morphing mimicked the effect of hormones on facial growth. It is becoming apparent that attractiveness, masculinity, growth, and sexual dimorphism are not all modeled by a small selection of images, nor are they collinear. The causes underlying each phenomenon differ and each phenomenon occupies a different level of biological explanation. In fact, each is complicated and poorly understood (Bruce & Young, 1998; Grumbach, 2000; Tanner, 1990). The morphed trajectory between images of men and women is but one description of sexual dimorphism and is subject to rejection (Meyer & Quong, 1999; O'Toole, Wenger, & Townsend, 2001). Other descriptions of dimorphism exist or can be developed.

Morphing models have a sculpture-like concreteness that gives them enduring appeal and credibility. It is frustrating that we cannot gain insight by simple attention to our perceptions of masculinity, because it is likely that these images are telling us something interesting. The outstanding problem is why the transformed face images appear masculine when masculinized. What makes them "masculine" and how is this different from perceived masculinity in real men's faces? I designed two studies to try to understand this. The first is an experiment to determine whether masculinity refers to how distinctly a face is recognized as a man or woman and whether morphed variation in masculine appearance changes these perceptions.

To determine why the masculinized images are different from unaltered images of masculine men it is necessary to employ a measurement model that is sensitive to variation in feature patterns that are perceived as masculine. Experiment 2 is designed to answer the question with such a model.

EXPERIMENT 2: A COMPUTATIONAL MODEL OF FACIAL MASCULINITY

> "Our computer graphic manipulations of the 'geometrical' differences between male and female face shapes generate stimuli that embody the 'psychological' meaning of masculinity and femininity." (Perrett & Penton-Voak, 1999, p.662).

Masculinization and masculinity have opposite effects on attractiveness, which is sufficient to suspect that transformed images are not accurate models of masculinity. It is true that masculinization of a face image causes it to be perceived as more masculine, however, it is difficult to determine what about masculinization is different from variance in masculinity. Perrett & Penton-Voak's (1999) hypothesis that the warped images are a psychological model of masculinity is difficult to validate, principally due to the complexity of facial variation. Visual inspection of the images can't tell us whether, or how, they misrepresent facial masculinity.

We need an efficient way to encode variation in facial structure, and the means to relate it to psychological judgments of masculinity. The encoding needs to be generalizable so that the masculinity of novel images of faces can be predicted. It is possible to test the warped-image-model hypotheses with a computational model of facial masculinity. Several researchers have constructed computational models to differentiate images of men's and women's faces (Cheng, O'Toole, & Abdi, 2001; Golomb, Lawrence, & Sejnowski 1991).

Testing the hypothesis that variance in the masculinity of transformed faces parallels between-sex differences in facial appearance is an extension of these models. That is, identify the information sufficient to distinguish unaltered faces of men and women, and then use this information to predict masculinity of transformed and unaltered images of men. If prediction of unaltered images *and* transformed images of men's facial masculinity is strong, this will provide supportive evidence for the hypothesis that transformed faces provide a model of within-sex variation in masculinity.

The hypotheses that within-sex masculinity parallels between-sex appearance and that masculine-appearance variation in warped faces is a model of men's facial masculinity can be tested using partial least squares regression (PLS). PLS explains variation in one data set in terms of a second corresponding set of data by finding factors that maximize the covariance between the two datasets (Geladi & Kowalski, 1986).

PLS is a successful analytic tool for datasets that have many more features than samples. For example, PLS has been used in studies of genetic expression to determine what base pair patterns within a gene distinguish unafflicted individuals from those who have a disorder suspected to be genetic in origin – in other words, PLS can work through datasets with large numbers of features to identify meaningful patterns – in this case, alleles of a gene (Huang & Pan, 2003).

Using PLS to explain psychological judgments given to faces is analogous to the gene expression problem. PLS is capable of extracting image-based components that distinguish images of men and women. It can also extract components that distinguish images of masculine and feminine men's faces. In this way, PLS can be used as a tool to reveal what features perceivers use to generate judgments about stimuli. Additionally, each location (that is, pixel) in a component can be assigned a statistic that describes how well that location predicts category membership or a perceptual judgment. PLS components and the pixel statistics can be visualized as images, and may have interpretable features.

PLS was used to answer two important questions. First, do the same factors that differentiate images of men and women also differentiate unaltered images of masculine and feminine men? Second, do the factors that differentiate men and women also differentiate images of men warped to be more or less masculine? I compared the image components PLS extracts that explain 1) the aspects of men and women's images in terms of sex of the individuals depicted, 2) variation in men's facial images in terms of their rated masculinity and femininity, and 3) variation in warped men's facial images in terms of their rated masculinity and femininity.

Method

Stimuli

Two sets of images were used for the simulations. The first set, called *unaltered*, consisted of images of 50 men and 50 women's faces. All faces were photographed with neutral expression, posed directly towards the camera. All images were part of a larger database of face images that were each previously rated for attractiveness. From this database we sampled equal numbers of faces from low, medium, and high attractiveness tertiles to represent a wide range of facial attractiveness in both men and women (alpha coefficients were .93 or higher). Masculinity ratings ranged 1.86-4.28 (M = 3.04) for men's faces and femininity ratings ranged 1.51-4.58 (M = 3.01) for women's faces.

Each image was digitized at 256 by 256 pixels in 8-bit grayscale. Face images were aligned manually using Adobe Photoshop, such that the circles described by the visible portions of the irises are aligned as well as possible. Transformations of the images to achieve eye alignment do not distort the aspect ratio of the face. Eye alignment is a common preprocessing step for automated face recognition algorithms (Hancock, Bruce, & Burton, 1998; O'Toole, Wenger, & Townsend, 2001).

The second set, called *transformed*, consists of the images used in the previous study. Eight original faces were transformed in 4 different ways (more average and more masculine, more average and more feminine, less average and more masculine, less average and more feminine). Additionally, the image transformations were performed at two different strengths (slight and moderate) to create 128 novel facial images. These images were the stimuli that the second group of participants rated in Study 1. The 64 additional faces were included because it was assumed it would help clarify the PLS results and so the numbers of transformed men's faces is more equivalent to the number of unaltered men's faces. Each of the 128 images was carefully aligned as described above.

Participants

A minimum of 40 undergraduates, relatively equal amounts of male and female raters, previously rated each of the 100 unaltered stimuli for masculinity/femininity on 5-point Likert scales Masculinity and femininity ratings were transformed so that they described a single scale on which men and women were separated. Scores were first z-transformed, and women's femininity z-scores were multiplied by -1 to produce "masculinity" scores. The maximum women's score was added to each z score representing men's facial masculinity. The transformed masculinity scores are intended to correspond to the hypothesis of Penton-Voak and Perrett (1999); within-sex differences parallel between-sex differences.

Analyses

The 100 images of unaltered faces (50 men) were resampled 100 times leaving out a single, different, image each time (that is, “jackknifed”). During each jackknife resample the 99 selected images were analyzed by PLS. The PLS analyses extracted several components that differentiated images of men and women that were selected in the resample. A linear perceptron used PLS component scores to classify face images as men or women. PLS components are the same size as each of the images and can be viewed as images. Each image in the PLS analysis was reflected off (that is, matrix multiplied) the components generated by the PLS analysis, generating a matrix that represented each face’s similarity to each of the components (that is, a 99 x *N* matrix, where *N* is the number of components used). This matrix was submitted to a two-layer linear perceptron. The perceptron has as many inputs as PLS components. It optimizes the weights on the inputs so it can best discriminate images of men and women.

The dropped out face is then reflected on the PLS components. This produces a small vector of activations to the trained perceptron’s input cells. The perceptron output is binary, indicating its decision of whether the image is of a man or woman. Whether or not the output cell is activated, however, depends on the summed *p* inputs multiplied by the optimized weight *w* on each input cell *i* plus the bias *b* used by the network (that is, y $=\sum_{i=1}^{n} p_i w_i + b$). If y exceeds threshold θ, the stimulus is assigned to category A, otherwise it is assigned to category B. The summed activation to the perceptron is continuous and is used as the predictor of masculinity.

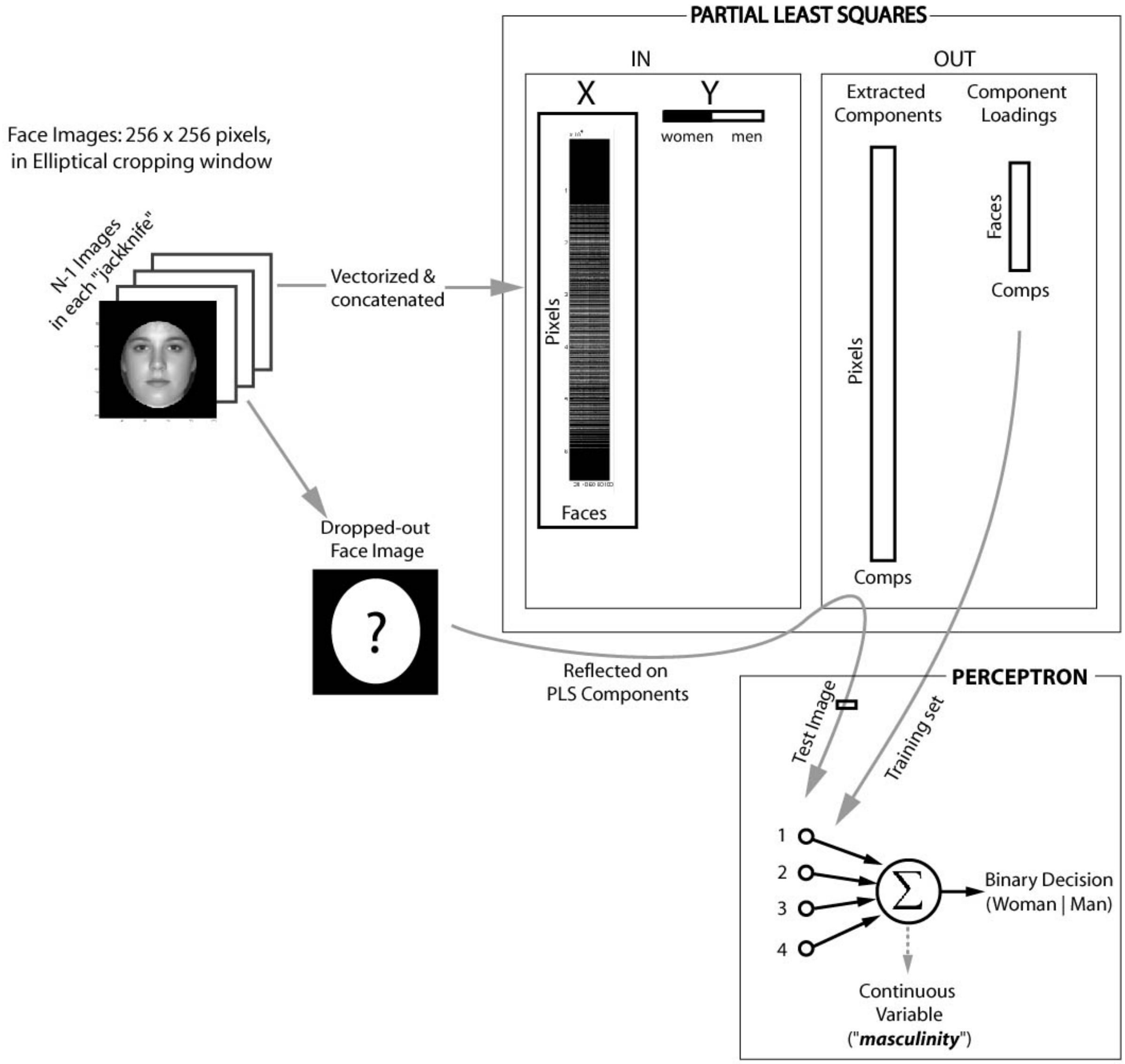

Figure 14: Modeling facial masculinity.

To determine whether within-sex masculinity of unaltered face images parallels between-sex differences, the perceptron network activations to each image were correlated with the mean masculinity/femininity rating of the unaltered images. Similarly, to determine whether variance in perceived masculinity of the transformed faces (due to morphed masculinization and feminization), parallels between-sex differences the transformed face images were submitted to the model trained to distinguish unaltered images of men and women.

A second type of analysis, using PLS alone, was conducted to determine the image features that contributed to judgments of sex and masculinity in the unaltered and transformed image sets. In this second analysis each pixel in an image was used to generate a separate prediction of the masculinity of the person depicted in the image. For each jackknife sample two PLS components were calculated that maximized the covariance between the image set and masculinity/femininity ratings made to those images. Each pixel in the dropped-out image was then multiplied by the "pixel" in the PLS component that matched its location in the image frame (that is, array multiplication), giving a prediction of the face's masculinity rating or sex. After each image was assessed in this manner there existed a distribution of predictions, representing each face, for each pixel in the image frame. Each pixel's distribution of predictions was correlated with the corresponding masculinity ratings given to the images for the analyses of masculinity, or a t statistic was calculated for how the PLS component predicted face sex at each pixel location. These operations should reveal which areas of faces are the best predictors of sex and masculinity.

Predictions

First, the features that discriminate men and women should moderately predict variation in perceived masculinity among unaltered men and women's faces, because within-sex variation parallels sexual dimorphism to some extent. Given the unusual findings with transformed faces in Experiment 1, the features that discriminate men and women should not predict the masculinity of transformed men's faces. Second, the pixel-by-pixel predictions of masculinity and sex should reveal the image features of unaltered and transformed images that distinguish faces by sex and masculinity.

Results

Shown in figure 15 are the first two PLS components that discriminate men’s and women’s faces. The component on the left has been photographically inverted from the original to show how its appearance is similar to the averaged female face; the original appearance of the component is displayed as the inset image. Four components were selected on the basis of examination of a scree plot for the PLS solution and because accuracy of prediction of the sex of transformed faces improved substantially to from 1 to 4 PLS components but not if more than 4 components were input to the perceptron. The components explained approximately 50%, 25%, 8%, and 7% of the variance in gender.

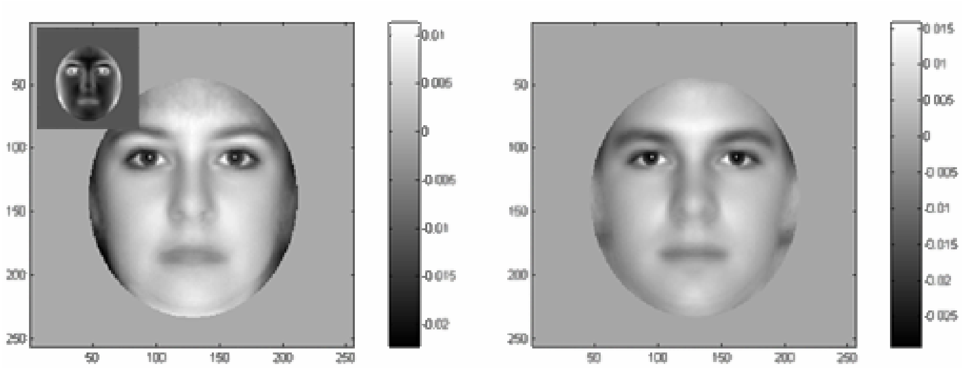

Figure 15. Two PLS components that differentiate images of men’s and women’s faces.

The computational model of facial gender was trained on unaltered faces and used to predict the masculinity ratings of unaltered men’s and women’s face images. Accuracy of classifying images of faces by sex was 92% (5 men and 3 women were misclassified). The unweighted PLS components correlated weakly with both men and women’s masculinity/femininity (from -.01 to .26). When the perceptron is trained to discriminate images of men and women, moderate prediction of men’s masculinity – but not women’s femininity – emerges. The summed network activations to the perceptron correlated +.4, $p < .01$, with men’s masculinity and -.06, *ns*, with women’s femininity.

The model was then used to predict masculinity ratings of transformed men's and women's face images. Accuracy of classifying transformed men and women by sex with information from images of unaltered faces was 91% with 4 PLS components (11 men and 0 women were misclassified). The summed network activation to the perceptron correlated .15, *ns*, with transformed men's masculinity and -.25, $p < .05$, with transformed women's femininity.

		Unaltered		Transformed	
		Men	Women	Men	Women
PLS Components	1	.05	-.26	.05	-.37**
	2	.21	.22	.06	.43**
	3	-.07	-.04	-.11	-.43**
	4	-.01	-.03	-.30*	.18
Summed Perceptron Activation		.40**	-.06	.15	-.25*

** p <.01, * p < .05

Table 1: Correlations of masculinity ratings and computational measures of facial sex.

These results indicate that variation in perceived masculinity in unaltered men's faces parallels between-sex differences to a moderate degree. By contrast, variation in perceived masculinity of transformed images of men does not parallel between-sex differences.

Another view of the differences between transformed and unaltered images of men's faces is to visualize the extent to which variance of the PLS component that explains gender correlates with masculinity judgments at each pixel location. The patterns of pixels that are strongly correlated with masculinity may be different in some way that sheds light on the problem (Figure 16).

The parts of faces that differentiate men and women form clustered patterns that appear to have some meaning. For example, the parts of the face that are the best cues to gender are the eyebrows, the forehead slightly above and between the eyebrows, and the periphery of the face (probably because of hair). The results agree with Russell's (2004) finding of pigmentation differences between men's and women's faces, especially that the area around the eyes is dimorphic. A large proportion of the pixel statistics are significant; t scores above 5 are significant after Bonferroni correction.

Figure 16: Pixel (*t*) statistics for differences between men's and women's faces.

Similarly, pixels that differentiate masculine and feminine images of transformed men form interpretable patterns. In contrast, however, the distribution of correlation values is bimodal and severely leptokurtotic. Almost every pixel is correlated .2 or -.2 with rated masculinity; the distribution is significantly different from 0. This striking result is not reflected by the pixel statistics for the PLS components that differentiate masculine and feminine images of unaltered men. The spatial and frequency distributions for unaltered images of men are more nearly random compared to the PLS pixel statistics for transformed images of men; the distribution is centered at zero and, with Bonferroni correction, only values beyond +/-.45 are significant (see Figure 17). A less conservative correction factor does not substantially alter the critical value for correlation significance.

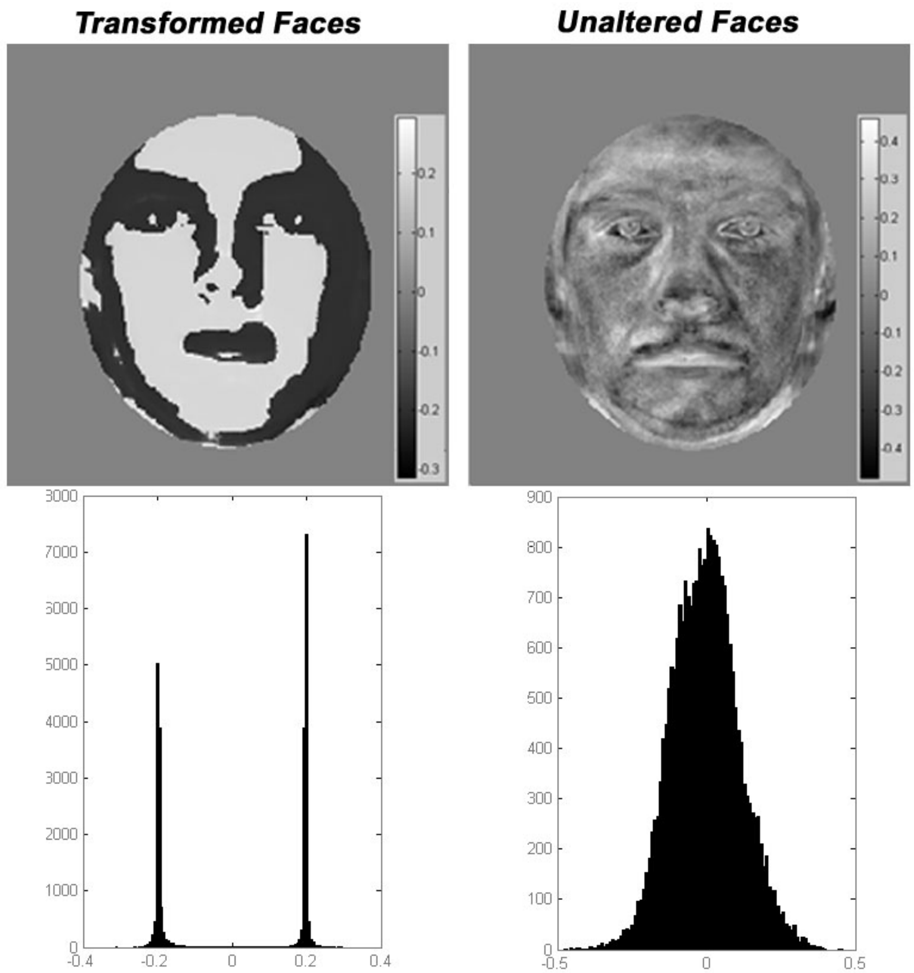

Figure 17: Comparison of correlations between transformed (left) and unaltered (right) men's masculinity and PLS prediction of sex for each pixel (first PLS component only).

Discussion

It is understandable that Penton-Voak and Perrett (1999) thought that their morphed images embodied masculinity. Their morphing transformations were designed to make an averaged man's face more or less similar to an averaged woman's face, and these transformations corresponded to changes in perceived masculinity. The model thus rests on reasonable assumptions; "masculinity" ought to refer to sex differences. Moreover, a model of sexual dimorphism that is based on global appearance differences between men and women – as the morphing model does – should capture some of the appearance differences that characterize masculinity in men's faces.

The finding that feminized men were more attractive than masculinized men (Perrett et al., 1998), however, was difficult to explain from contemporary evolutionary psychology theories. "Good gene" theories favored by evolutionary psychologists predicted that masculine men should be perceived as more attractive. Experiment 1 demonstrated that feminization and masculinity can both simultaneously make men's faces more attractive. Thus, feminization and femininity can not be equivalent. The results of these simulations confirm this; warped images do not embody facial masculinity and are not a distillate of sexual dimorphism.

The statistics for each pixel provide a necessary clue to explain why the warped-image-model is not an accurate model of masculinity. In the transformed images, almost every pixel is correlated with masculinity to a small degree, but this is not true for the unaltered images. The discrepancy is probably not attributable to differences in the rules by which perceivers view and rate the images; participants probably apply the same "rules" to judge the masculinity of transformed and unaltered images. Rather, the discrepancy can be explained by the different processes for masculinizing men's faces. In image morphing, every pixel in the image is changed a small amount to make it more or less "masculine." This is not true for the natural masculinization process achieved through facial growth. Growth is messy; the timing, amount, and pattern of growth differs between individuals and growth involves nonlinearities that cannot be replicated by conventional morphing algorithms.

It is possible that if averageness is independent of sexual dimorphism, that the relevant comparison might be to a "sexless" prototype face. If averageness indexes how fast a face is recognized as a face, then facial sex may be irrelevant to the averageness hypothesis. Thus, feminizing a man's face could change it to be more like a sexless prototype. In Study 3, in which averageness is modeled, the hypothesis of whether comparisons should be to sex-typed averaged faces was tested by determining whether similarity to sex-congruent or combined-sex averaged faces better predicts the attractiveness of an unaltered face image.

STUDY 3: MODELING THE AVERAGENESS HYPOTHESIS

It has been shown that when faces are combined by pixel averaging that they become more attractive (Langlois & Roggman, 1990) and that when individual faces are morphed to be more similar to an averaged face they become more attractive (Lee, Byatt, & Rhodes, 2000; Rhodes, Sumich, & Byatt, 1999; O'Toole, Price, Vetter, Bartlett, & Blanz, 1998, and Experiment 1). To test the averageness hypothesis it is critical to determine whether attractive faces are more like averaged faces than are unattractive faces; it has not yet been established whether this is the case.

This dissertation has reviewed research in which faces are measured as if they are uncomplicated objects (e.g., Cunningham, 1986; Grammer & Thornhill, 1994), which is not acceptable because feature distance measures are insufficient to capture within-category variation. Facial measurements are rarely developed to represent how people might perceive faces.

It is critical to establish that cognitive modeling and facial measurement are compatible. In fact, the congruence between modeling facial cognition and measuring faces is exploitable. Models of facial cognition can be used to construct biologically plausible measurement methods. According to theories of knowledge representation, perceivers internally represent the outside world. Thus, a model of the world exists in the perceiver's mind. Therefore modeling the perceiver's representing world transitively models the real world (Palmer, 1975). Moreover, Edelman (1998a) argues that representation is essentially measurement followed by dimensionality reduction. The aim of dimensionality reduction is to account for redundancy, or information, in the stimulus input and to ignore sources of variance that are not informative.

Computational approaches used in cognitive modeling reduce a large number of image features to a much smaller number of components, but this is very different from the discarding of information that inevitably occurs with facial feature distance measurements. When dimensionality is reduced by accounting for redundancy in the data, knowledge is generated. Such knowledge allows detection of suspicious coincidences in the data (Barlow, 1989; Bartlett, 2001). Organisms identify and account for redundancy in perceptual input so that they can later quickly detect relevant and meaningful stimuli. With respect to faces, the relevant redundancies are the salient axes of within-category variation (for example, round vs. long, man vs. woman).

Study 3 both quantifies a face's similarity to an average face and simulates the process of prototype formation predicted by the averageness hypothesis. The test is a computational model simulation that forms averages of men's and women's face images and extracts sources of variance in the images. The similarity of new face images to the averaged faces of men and women is calculated with reference to the sources of variance in the set of images. The similarity of new faces to the sex-congruent averaged faces is their predicted "attractiveness," that is compared with human attractiveness ratings of the same faces. The process is repeated by random sampling from a database of face images. The resampling process establishes the reliability of the association between attractiveness and averageness.

Principal Components Analysis

Researchers in face perception have used several different recognition algorithms to model face representation and recognition. The most popular, due to its ease of implementation and good performance, is Singular Value Decomposition or Principal Components Analysis (PCA). It is the most frequently-used baseline for comparison with other computational models (Cottrell, Dailey, Padgett, & Adolphs, 2001).

PCA is calculated on the face images (that is, every pixel within each image is input to the PCA) and is mathematically equivalent to a type of neural network called an autoassociator, which gives a rotated PCA solution. PCA is a data reduction method that represents a large number of objects using a substantially smaller amount of data. In the case of faces, PCA extracts a set of basis images called eigenvectors (or eigenfaces) from a set of images. These eigenvectors are the principal axes of configural variation in the set of face images.

PCA identifies orthogonal, linear, factors in a data set. From a set of N images, PCA extracts N-1 principal components. Principal components are also called eigenvectors, eigenfaces, or basis images because of what they are mathematically, how they appear, and how they are used, respectively. The components describe configural variation among facial images, satisfying Young & Bruce's suggestion that face representations "capture the differences between the faces we encounter" (1991, p. 13), as well as Barlow's (1989) idea that accounting for regularities in the stimulus input provides information.

Each basis image is the same size and shape as one of the original images, and each of the original images can be represented perfectly as a vector of weights for each of the basis images. In other words, each face image that was analyzed by PCA has a score on the first, second, third, etc. basis images that identifies how much of that image to use in order to reconstruct the face.

A benefit of this form of compression is that it can encode face images that were not used to construct the principal components. Novel faces are represented, though imperfectly, by a vector of weights that refer to the eigenvectors. PCA basis images contain configural information about faces that generalize to many possible faces. Thus it is useful for automated face recognition, because PCA can be used to both locate face-like stimuli in images and compare them to previously seen faces.

The basis images reveal interesting things about face structure. For example, if men's and women's faces are included in a PCA analysis, the first principal component (or basis image) generally is an image of the difference between men's and women's faces (O'Toole, Vetter, Troje, & Bülthoff, 1997) (see image #1 in Figure 18 below). If this basis image is added or subtracted from the average face, the transformed image is either masculine or feminine.

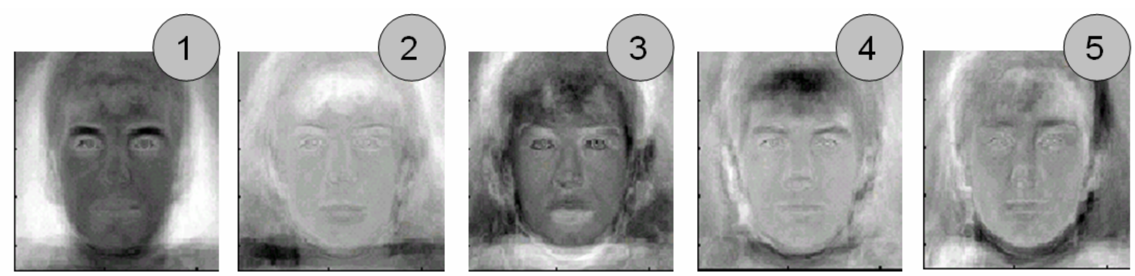

Figure 18: The first 5 eigenvectors of the facial images used in Study 3.

PCA is used to model face perception, representation, and recognition because of its general plausibility and its congruence with theories of face perception. There are several properties of the model that match theories of face perception. First, the general structure of PCA - image, weights, and basis images - is a simplified model of the retinal representation, neural activation, and cortical representation of faces, respectively

(Cottrell, Dailey, Padgett, & Adolphs, 2001; O'Toole, Wenger, & Townsend, 2001). Second, the information PCA uses – all of the image pixels – preserves information thought to be vital for face recognition – shading and configuration (Davies, Ellis, & Shepherd, 1978; Thompson, 1980).

Third, the PCA representation of faces is compatible with face-space models of mental representation; the weights on the eigenvectors that are used to reconstruct faces in the PCA describe a high-dimensional space in which faces are represented as single points and the distance between points indicates similarity. These are the properties of the face-space metaphor of face representation in which the average face is at the center of the space. Furthermore, I discussed earlier how feature-based models suffer from several problems, one of the most serious being that feature selection is an idiosyncratic process subject to experimenter bias. The PCA model does not suffer from this problem; all image features (pixels) are used.

The general similarities of the face-space model and the PCA computational model have been used to make predictions about behavioral data in tests of the face-space model and theories of representation. For example, the "other-race" phenomenon is the effect that an observer is better at remembering new faces of his or her own race than that of a different race. The effect is hypothesized to exist because all faces are very similar to each other and distinguishing among them requires sensitivity to slight configural differences that is acquired with extensive experience (Valentine, Chiroro, & Dixon, 1995). People might be experts at distinguishing new faces but may not be sensitive to sources of facial variation individuating people of a different race. The other-race effect has been modeled using PCA, lending credence to the use of PCA as a model of human face perception (O'Toole, Deffenbacher, Valentin, & Abdi, 1994).

In this study I will use PCA as a method to measure the averageness of faces and determine whether faces that are more similar to an averaged face are also more attractive. Compared to previous studies testing the averageness hypothesis, this measurement method is a more useful model of face recognition and discards no features (Grammer & Thornhill, 1994; Jones, 1996; Pollard, Shepherd, & Shepherd, 1999).

Independent Components Analysis

Although PCA is often used to model face perception, there are several alternative architectures. Independent Components Analysis (ICA) is an alternative that differs from PCA in that the components that describe variation in the image set are allowed to be correlated. PCA describes second-order image statistics, which refer to sources of contrast in the image set. ICA describes second- and third-order image statistics. Third-order statistics are relations among groups of pixels, which include small features such as lines and curves that recur in the image set (Bartlett, 2001).

ICA components formed from analysis of geometrically unaligned stimuli, such as images of natural scenes, look like neural receptive fields (Bell & Sejnowski, 1997). ICA components formed from analysis of aligned images of faces include what appear to be local features that correspond to some nameable parts of the face, such as "chin," "iris," "upper lip" (see Figure 19, below).

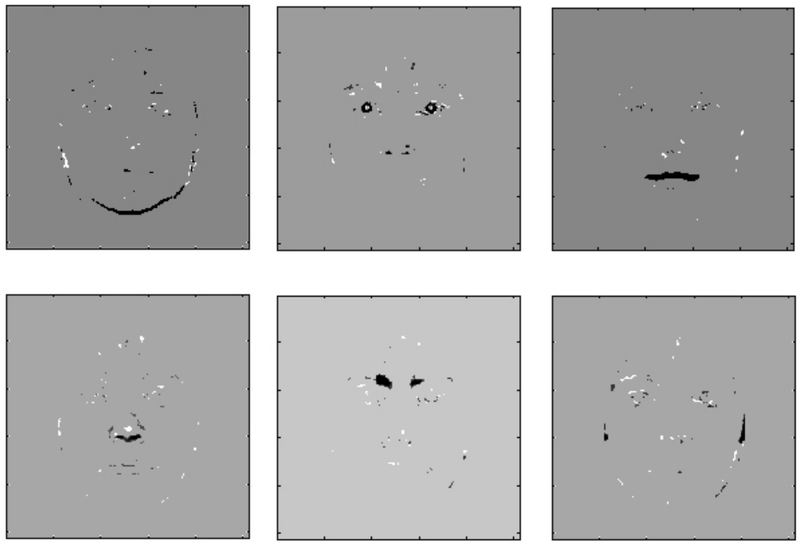

Figure 19: Pixel relations represented by several ICs calculated from the images used in Study 3.

Bartlett (2001) found that ICA is in some conditions superior to PCA for face recognition. ICA provides a means of face representation more sparse than PCA, which permits it to better account for redundancy in images of faces, maximizing information transfer. Bartlett used a database of face images that had images of individuals taken on different days and with different expressions to determine how well PCA and ICA performed recognition. ICA correctly identified a significantly greater percentage of individuals whose pictures were taken on different days. For images of people taken on the same day but who assumed a different expression PCA and ICA performed approximately equally.

Partial Least Squares

Partial Least Squares (PLS) was described in Experiment 2. PLS extracts components from one dataset that explain variance in terms of a second set of corresponding data. In the case of image data, PLS is a useful tool; PLS can identify the features that distinguish images, based upon human judgments of the images. PLS will be used as a front end to two different neural networks: First, a linear perceptron; second, a nonlinear backpropagation network.

In the case of attractiveness ratings, PLS should be able to extract image-based components that distinguish face images by their attractiveness. PLS will be used to answer two important questions: First, if PLS is able to encode information about how humans formulate attractiveness judgments of faces, does this inform us about the averageness and sexual dimorphism hypotheses? Moreover, how well is attractiveness explained by a few linear components that are combined linearly or nonlinearly? The extent to which attractiveness judgments can be effectively captured by a model using PLS should give some indication as to what facial attractiveness *is* - whether it is encodeable from the facial image (and generalizable to novel images). In this way this model should provide a benchmark for hypothesis-driven computational approaches, such as the PCA and ICA analyses.

Method

Stimuli

The images, same as the unaltered faces used in Study 2, were 100 faces (50 men) of college-aged Caucasian students. Equal numbers of faces from low, medium, and high attractiveness tertiles were sampled to represent suitable variation of facial attractiveness. Mean attractiveness ratings ranged from 1.03-4.06 ($M = 2.39$) for men's faces and 1.33-4.19 ($M = 2.53$) for women's faces. Each image was 256 by 256 pixels in 8-bit grayscale. Face images were aligned manually using Adobe Photoshop, such that the circles described by the visible portions of the irises are aligned as well as possible.

Portions of each image were also submitted to the computational models after an elliptical cropping window was applied to the image set. Only pixels within the window were analyzed to control for the effects of variation in amount and style of hair, which is relatively unstructured compared to faces. For each image the cropping window was the same size, every un-cropped pixel location in each image was represented in every other image. As the frame had to be the same size for every image, the proportions were chosen to be an optimal fit for men's and women's faces (see Appendix A for details).

Participants

A minimum of 40 undergraduates, relatively equal amounts of male and female raters, previously rated each of the 100 stimuli for attractiveness on 5-point Likert scales (alphas = .93 or higher).

Analyses and Predictions – PCA and ICA

If attractiveness of individual faces and similarity to an averaged face are positively and significantly correlated, this will support the averageness hypothesis. Similarity to averaged faces and attractiveness were correlated in several different conditions to determine whether similarity to sex-appropriate averaged faces, opposite-sex averaged faces, and combined-sex averaged faces predicted attractiveness ratings.

To test the hypothesis that attractive faces are average in appearance, face images were jackknifed. *N*-1 images were submitted *N* times to PCA or ICA algorithms that identify sources of variance in the images. Each sample of images was divided into three sets for analysis: men only, women only, and both men's and women's faces. In each analysis, the set of sampled faces are combined to construct the average male, female, or combined sex face.

The unsampled face is matrix multiplied by the extracted components, which "projects" them into the PCA or ICA face space created from the sampled faces. This projection gives the face image a weight on each component. The similarity, or distance, to the averaged face is calculated from these weights. As three different spaces are calculated for each jackknife resample, the unsampled face is compared to the female averaged face calculated from the women's faces in the analysis, the male averaged face calculated from the men's face images in the analysis, and the combined male+female averaged face calculated from the men's and women's face images in the analysis. The averaged face images will be referred to as X_F (female), X_M (male), and X_{MF} (female+male) to distinguish them from the images of specific men and women.

For PCA analyses, similarity was calculated using two different similarity metrics, Euclidean and Mahalanobis. The Euclidean metric weights the lower-order components more strongly. The first components extracted have been shown to correspond to information that is useful for making semantic judgments of faces (Valentin & Abdi, 1996). On the other hand, as the Mahalanobis distance equalizes the component space and places relatively more value on the individuating information that tends to be represented in the components with smaller eigenvalues. The Mahalanobis distance metric has been shown to outperform the Euclidean metric in automated face recognition simulations (Burton, Miller, Bruce, Hancock, & Henderson, 2001; Moon & Phillips, 2001). The superiority for the Mahalanobis metric is likely because facial recognition functions better when higher order components are available for making recognition decisions. It is also predicted that attractiveness judgments will be best approximated (through computing similarity to averaged faces) when the information by which similarity is calculated is also better suited for face recognition. Thus, the

Mahalanobis metric should produce higher correlations with attractiveness than the Euclidean metric.

Euclidean distance is calculated as the square root of the summed and squared component weights. This is the length of the face images' vector in Euclidean space. The Mahalanobis distance is calculated by the Euclidean formula after the weight space has been normalized by the square root of each eigenvector's eigenvalue, so that each component accounts for the same amount of variance.

Analyses and Predictions – PLS

The images were again jackknifed, and PLS extracted components that maximized the covariance between the image pixel values and the mean attractiveness ratings given to the faces. The attractiveness of the dropped-out faces was estimated by matrix multiplication of the images by the extracted components, which generates scalars that are predictions of the faces' attractiveness scores. These predictions were combined by a linear perceptron or backpropagation network that weighted the influence of each component in order to optimize prediction of attractiveness of the faces in the jackknife sample.

Single-unit perceptrons are designed to make binary decisions (for example, "man" or "woman"). Feedback to a perceptron is, in the classical case, also binary – during training the network is told whether its decisions are correct or incorrect and it adjusts the weights on its inputs accordingly. Here the perceptron was trained to emulate a continuous variable, which is called the Widrow-Hoff or delta rule (Hagan, Demuth, & Beale, 1996). The network activation (that is, the inputs multiplied by the input weights plus the bias) is used as the perceptron's measure of attractiveness.

The multilayer backpropagation network overcomes the limitations of perceptrons, which can only solve linearly separable problems. Multilayer networks can approximate any linear or nonlinear continuous function (Hagan, Demuth, & Beale, 1996). The backpropagation network had one hidden layer, which was tested with one, two, or three units.

In each jackknife, the PLS analysis extracted components for 1) men's faces only, 2) women's faces only, and 3) both men's and women's faces that related variation in the images to the images' attractiveness ratings. These components were submitted to a 1) linear perceptron or a 2) multilayer backpropagation network that were trained to weight each PLS component to optimize the prediction of attractiveness of faces that were selected for the PLS analysis. Within each group of faces, one face image was excluded from both the PLS and network training.

Results

PCA

These analyses were designed to answer two questions: First, whether PCA-derived similarity to average faces predict face attractiveness ratings, a result that would support the experimental hypothesis that faces that are more average are more attractive. Second, given that the similarity scores predict attractiveness ratings, whether the Euclidean or Mahalanobis metric explains more variance in attractiveness ratings.

Cohen & Cohen's (1983, p.57) formula that determines the significance of the difference between dependent correlations was used to establish whether the Euclidean or Mahalanobis correlations with attractiveness were significantly different. The Euclidean and Mahalanobis correlation coefficients are dependent because the attractiveness ratings refer to the same set of faces. Table 2 shows that similarity of a face image to an averaged face weakly or moderately predicts its attractiveness. It is also apparent that there is an advantage to using the Mahalanobis distance metric over the Euclidean metric for predicting women's, but not men's, attractiveness.

		Correlations with Distance to Average and Attractiveness of Face Images		
		Male + Female	Male	Female
Images	Metric	Average	Average	Average
Men	Euclidean	.32, p<.05	.34, p<.05	.26, p=.07
	Mahalanobis	.32, p<.05	.31, p<.05	.24, p=.09
Women	Euclidean	.22, p=.12	.23, p=.11	.09, p=.53
	Mahalanobis	.42, p<.01	.28, p=.05	.25, p=.08
Men & Women	Euclidean	.27, p<.01	.28, p<.01	.17, p=.09
	Mahalanobis	.37, p<.0001	.30, p<.01	.24, p<.05

Notes: Gray shading indicates that paired Euclidean and Mahalanobis correlations within a cell are significantly different.

Table 2: Correlations of similarity to averaged face and attractiveness of men and women's images.

It appears that using similarity to a combined male+female averaged face better predicts women's facial attractiveness than using similarity to a female averaged face. For men, similarity to a combined male+female averaged face predicts men's facial attractiveness as well as does similarity to a male averaged face.

In a second set of analyses I subsampled the face images and calculated their similarity to male, female, and male+female averaged faces. Subsampling is a type of resampling in which a portion of the images – in this case approximately 25 men and 25 women – are selected for analysis, and the process is repeated many times. I subsampled the face images 77 times. Efron & Tibshirani (1993, p. 52) state that 25 resamples is "usually informative" and 50 resamples "gives a good estimate of" the standard error of the statistic being resampled.

The subsampling results were similar to the jackknifed correlations shown above. The subsampling allows construction of an empirical sampling distribution of correlations between similarity to an averaged face and attractiveness. The distributions are summarized in Figure 20, which shows the mean and standard deviation of correlations between similarity to an averaged face and attractiveness as a circle plus error bars (Mahalanobis metric). The mean of the Euclidean metric is shown as a small square.

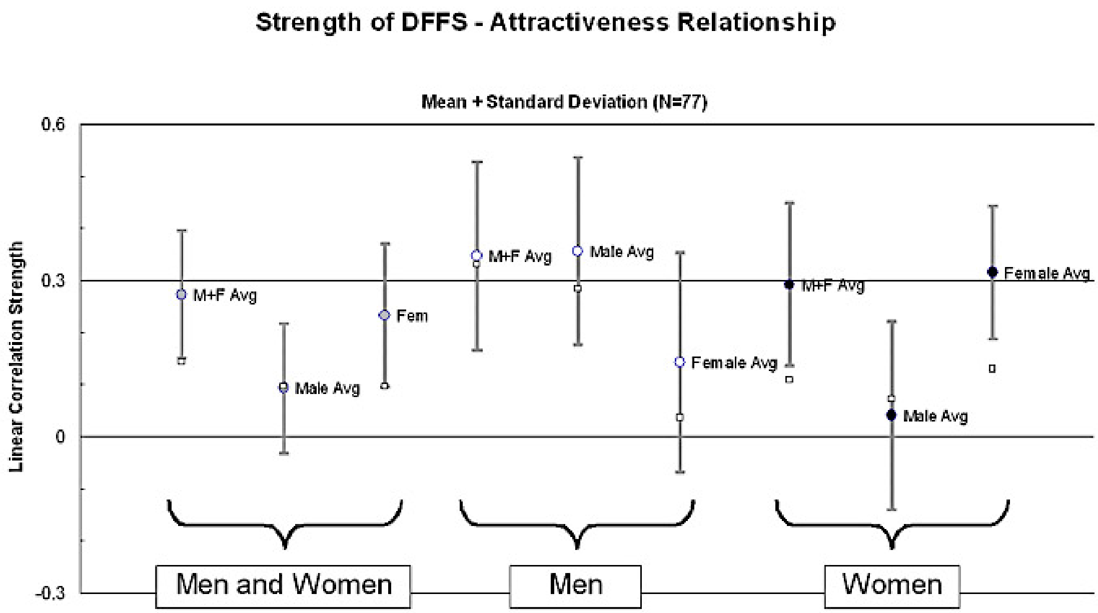

Figure 20: Summary of subsampled correlation distributions.

ICA

The images of men and women's faces were jackknifed and submitted to Independent Components Analysis (ICA). In ICA space, the similarity of the dropped-out face to X_{MF}, X_M, and X_F was calculated. The results were of similar magnitude to those obtained with PCA. Again, there are some unexpected results such as similarity to the combined male+female averaged face (X_{MF}) and similarity to the male average (X_M) predicting women's attractiveness to a greater degree than similarity to the female average (X_F).

	MF Average	M Average	F Average
Men	.25, *p*=.08	.29*	.13, *p*=.2
Women	.31*	.31*	.28*
Both	.28**	.27**	.21*

**$p < .01$, *$p < .05$

Table 3: Correlations of ICA-measured similarity to averaged face and attractiveness.

The ICA-derived predictions of attractiveness ratings were not significantly different from the PCA-derived predictions using Cohen & Cohen's (1983) formula to determine whether two dependent correlations are significantly different. The PCA derived predictions, however, are generally higher than those made by ICA.

PLS

Facial attractiveness was modeled using partial least squares (PLS) as a front-end to two different neural networks. The first was a linear perceptron. The second was a backpropagation network that can approximate nonlinearities in the PLS weight space.

PLS was used to extract components from the face images that covaried with the attractiveness ratings made to the faces. The networks then were trained to optimally combine the PLS components to produce predictions of the images' attractiveness ratings. Finally, the trained networks were used to predict the attractiveness rating of an image which had not been used for neural net training nor had been included in the PLS analysis.

The results of the perceptron network were excellent in comparison to existing methods (see Table 4). Quite surprisingly, the backpropagation network made nearly perfect predictions of attractiveness ratings humans had made to images of both men's and women's faces. The results depended on which images were analyzed with PLS and which images were used to train the networks.

PLS Components: Men and Women's Attractiveness

		Women		Men	
		Attract	Masc	Attract	Masc
PLS Components	1	.38	.31	.13	.14
	2	-.13	-.10	.14	.01
	3	-.05	.09	-.01	-.17
	4	.09	.01	-.04	.11
Perceptron Training	Both	-.64	-.61	-.55	-.11
	Women	-.50	-.49	.05	.21
	Men	.65	.56	-.67	-.30

PLS Components: Men's Attractiveness

		Women		Men	
		Attract	Masc	Attract	Masc
PLS Components	1	-.01	.14	.14	-.07
	2	.32	.24	.25	.26
	3	-.28	-.06	-.03	-.07
	4	-.14	-.06	.004	-.07
Perceptron Training	Both	-.87	-.75	.62	.14
	Women	-.56	-.53	.16	.03
	Men	-.32	-.21	-.50	-.25

PLS Components: Women's Attractiveness

		Women		Men	
		Attract	Masc	Attract	Masc
PLS Components	1	.35	.32	-.01	-.004
	2	.21	.14	.03	.04
	3	-.01	.10	.17	.19
	4	-.003	-.07	-.19	-.27
Perceptron Training	Both	-.93	-.77	-.85	-.21
	Women	-.75	-.64	.07	.08
	Men	-.62	-.52	-.72	-.19

Table 4: Correlations between attractiveness and PLS or perceptron output.

Table 4 shows some very interesting patterns. First, the PLS correlations with attractiveness (disregarding the sign of the correlation, average $r = .13$) tend to be smaller than the perceptron network model correlations with rated attractiveness (disregarding the sign, average $r = .67$).

Table 4 shows that PLS components derived from only men's or only women's faces can be used in a perceptron to greater predictive efficacy than can PLS components derived from both men's and women's face images. It appears that the PLS components derived from images of both men and women, and their attractiveness ratings, are a relatively poor compromise. Moreover, the best models use PLS components derived from either men's or women's images and use perceptron training on images of both men and women.

Table 4 also shows that when PLS components derived from men are trained on both men and women, the perceptron activation predicts attractiveness strongly but in the opposite direction for men and women. By contrast, PLS components derived from images of women trained on both men and women predict men's and women's attractiveness in the same direction. Moreover, if the correlations for men (.62 and -.85) are squared and summed the result is close to 1.

The backpropagation network results show a very similar pattern, except that it is able to predict attractiveness ratings of faces, not trained or in the PLS, perfectly (correlations of +/-1), once outliers are removed. For approximately 98% of the face images the backpropagation network predicts their attractiveness perfectly, but the few remaining images are very poorly predicted (that is, the outliers). The network is able to maintain production of predictions that correlate 1 to untrained faces with one cell in one hidden layer.

Discussion

The results using PCA and ICA support the averageness hypothesis of facial attractiveness. With two different models, it was demonstrated that, in some conditions, images that are more similar to averaged faces are more attractive.

These measurement methods have been demonstrated to be useful as automated face recognition. That a reasonable prediction of attractiveness falls out of the unprocessed PCA and ICA weight space (without much finessing) is good support for averageness theory; these are but two possible computational encodings of the hypothesis. The model providing a test of the averageness hypothesis involved analyzing a set of images to find the (image-based) factors that varied among the images. There are other possible schemes for developing models to test the averageness hypothesis, such as determining whether the factors that differentiate faces and non-faces also differentiate attractive and unattractive faces. The models can be designed to epitomize assumptions of theories of categorization and learning.

Two different averageness measurement models have been proposed previously, one similar to the PCA model, and one that relied on feature distance measurements. O'Toole et al. (1998) showed that a PCA-based representation could account for a portion of variance of men's, but not women's, facial attractiveness. The model O'Toole et al. used differed from the PCA model that used the Mahalanobis metric, but was identical to the PCA model that used the Euclidean metric. O'Toole et al. did not subtract the averaged face from each image before PCA analysis, and used the weight on the first eigenvector as the measure of attractiveness. When the images are not preprocessed by subtracting the average of the images from each image, the first eigenvector is the average of the images. Thus, the weight of each face on this eigenvector is the amount of the averaged face required to construct the face. This is a reasonable measure of averageness; it is a possible encoding for “activation of the prototype receptive field.” The weight on the first eigenvector of images that have not been mean-centered is mathematically equivalent to the Euclidean distance to the center of the space of images that have been mean-centered. Thus, Study 3 replicated O’Toole et al.; the Euclidean metric predicted men’s, but not women’s, attractiveness.

The Mahalanobis metric explained more variance in women’s attractiveness than did the Euclidean metric. For men, the two metrics explained approximately the same amount of variance. Subsampling analyses revealed that the Mahalanobis metric explained more variance in men’s attractiveness than the Euclidean metric. The implication of these findings is that for both men and women, attractiveness judgments are better approximated by facial information which is also better suited for facial recognition. The Euclidean distance tends to place importance on information common to the images in the set, which better represents categorical information such as gender, whereas Mahalanobis distance weights all components in the PCA solution equally, which emphasizes the individuating information present in higher-ordered eigenvectors (Valentin, Abdi, Edelman, & O’Toole, 1997). It may be that women’s attractiveness is more dependent than men on the higher spatial frequencies, that can reveal, for example, smoothness of skin.

It is curious that ICA did not predict attractiveness ratings better than did PCA. Bartlett (2001) showed ICA to be superior to PCA for automated face recognition. Bartlett evaluated two different ICA models, one in which she defined an image space in which pixels were vectors and a second in which she defined a pixel space in which images were vectors. Study 3 used the second ICA architecture; the first is difficult to implement without using low-resolution images (Bartlett’s face images were 50 by 60 pixels). Bartlett found optimal performance through a combination of the two ICA architectures. Moreover, ICA performance with a single ICA architecture did not significantly outperform PCA if the individual to be matched was photographed on the same day but with a different expression. Thus, it is possible that a different ICA architecture could predict more variance in attractiveness ratings than does PCA, but the current results are not unexpected given Bartlett's findings. A replication and extension of the analyses presented here could more deeply explore possible ICA architectures.

In the literature there has been speculation about subcategory prototypes and how they might be appropriate for averageness theory (e.g., O'Toole, Wenger, & Townsend, 2001; Valentine, Chiroro, & Dixon, 1995). This could lead to redundancy between the averageness and sexual dimorphism hypotheses, which may or may not be problematic because if sex must be accounted for by averageness then its theoretical domain overlaps with sexual dimorphism. The findings from Study 3 suggest otherwise; the hypotheses should be independent explanations of facial attractiveness as similarity to combined male and female averaged faces predicted attractiveness as well as (for men), or better than (for women), similarity to same-sex averaged faces.

Attractiveness does not have to be a gendered concept; we are very sensitive to gender as a social cue but it is possible that we overlook many similarities between men's and women's faces. Moreover, "face" is one of the most important basic-level categories, so it is possible that the speed at which the judgment of whether the stimulus a face or not is critical to the initial stage of attraction.

PLS & Neural Network Models

The results raise many interesting questions, but they are also informative for examining how successful different training schemes are for predicting men's and women's facial attractiveness. Men & women's attractiveness seems largely described by a single combination of factors that PLS is able to extract by maximizing the covariance between images of women's faces and their corresponding mean attractiveness ratings. The perceptron can find a combination of these PLS components that generalizes to novel faces, explaining very large portions of variance in both men's and women's facial attractiveness (see Table 4). This suggests that a common factor underlies much of both men's and women's facial attractiveness.

Men's attractiveness, additionally, is predicted by a second factor, a combination of PLS components derived from images of men's faces and their attractiveness ratings. Scoring high on this factor indicates high attractiveness if the image is of a man but low attractiveness if the image is of a woman. With some speculation, it is possible that the first factor, explaining both men and women's attractiveness in the same way, could be similar to averageness. The second factor could be sexual dimorphism, as it positively predicts men's attractiveness and negatively predicts womens' attractiveness.

The most surprising finding of Study 3 is that the rules for deriving facial attractiveness can be learned and mimicked almost precisely by a relatively simple mechanism. This is especially striking because the predictions' accuracy exceeds typical interrater agreement found in studies of facial attractiveness, which have median interrater correlations of approximately +.7 (although estimated population reliabilities are ~.9) (Langlois et al., 2000). If attractiveness is so easily encoded, why do people not agree with each other more strongly?

Although the model is atheoretical (it simply mimics human judgment) it allows us to answer some interesting questions. First, we ask it to rate photographs of men for attractiveness as if the images were women, a task we expect a participant to perform with much difficulty. Second, we've seen that observing how performance varies with the conditions under which the network is trained helps us understand attractiveness. Third, it will be quite informative to determine where the model breaks down. Some ideas for further research are investigating “other-race” effects, and whether obscuring parts of the image result in decrements in performance that match human decrements. Further, if mean attractiveness judgments can be mimicked, perhaps so too can the preferences of an individual (it may be possible to thereby have an individual rate him or herself for attractiveness “objectively”). Last, it defines a standard for models of attractiveness in terms of performance and complexity that should only be compromised for increased biological plausibility.

Observing that these complex judgments can be encoded lets us speculate about the nature of attractiveness. We are fairly well-conditioned to make attractiveness judgments – 1) we are accustomed to using unidimensional scales to judge many things, 2) Many individuals are accustomed to thinking of attractiveness as an objective *property* of people, and 3) We are accustomed to thinking of attractiveness socially – we compare our judgments of attractiveness to those that others make. Attractiveness is as much embodied within a culture as within each person in the culture, perhaps more so. Developing a sense of attractiveness could very well proceed along two diverging paths, one personal and one cultural. I predict that they both begin with perceptual learning of faces.

In general, PLS may be useful for understanding how people make judgments to other complex stimuli. The success in prediction of human judgments to images lies in the extremely efficient dimensionality reduction PLS affords. Being able to represent each image as 4 numbers, without sacrificing fidelity with respect to the variable of interest, makes it relatively easy for the perceptron or backpropagation networks to find a solution from which useful generalizations can be made. That the components themselves explain variance in the variable of interest is the fundamental advantage of the model. It is not surprising that the network makes good predictions. What is surprising is how easily the network exactly reproduces mean attractiveness ratings of images on which it has not been trained.

Chapter 4: *General Discussion*

In these studies I have tried to explain facial attractiveness in terms of two different hypotheses, averageness and sexual dimorphism. The results of the experiments and simulations supported both hypotheses and sharpen our understanding of the phenomena. I will discuss how the results relate to each of the hypotheses.

IMPLICATIONS FOR THE MAJOR HYPOTHESES

Sexual Dimorphism

Sexual dimorphism refers to the difference between men and women, but the sexual dimorphism hypothesis of facial attractiveness is concerned with within-sex variation in facial appearance that reflects sex differences. Thus, the apparent masculinity/femininity of facial appearance is the quality that the sexual dimorphism hypothesis proposes explains facial attractiveness.

Proponents of the sexual dimorphism hypothesis believe that sexual dimorphism indexes mate quality. The degree to which a man is "masculine," or different from women, may reflect the functioning of his immune system (Thornhill & Møller, 1997). According to the hypothesis the amount of pubertal testosterone, to which a man's body was subjected, corresponds to both the degree of masculinization of his face and to the degree of immune system depression (Johnston et al., 2001). Under pathogenic infection, the body can either fight infection or develop a masculine face; presumably, the successful strategy is to curtail hormone production and boost immunity. Therefore, it is presumed that the body's natural ability to ward disease is the crucial factor that determines whether a boy grows into a masculine man.

A version of the hypothesis also provides for flexibility of preference direction so that a woman prefers feminine men when she is not fertile, but prefers masculine men when she is fertile (Perrett et al., 1998). Perrett et al. explain that feminine-appearing men should be more caring and devoted fathers, so women's preferences describe a rhythm in which masculine men provide sperm and feminine men provide care.

Studies using unaltered images of men's faces showed that faces rated more masculine were more attractive than faces rated more feminine (Brown, Cash, & Noles, 1986; O'Toole et al., 1998). By contrast, studies using morphed images of men's faces showed that feminine men's faces were more attractive than masculine men's faces (Perrett et al., 1998; Rhodes, Hickford, & Jeffery, 2000). It was not possible to determine whether the discrepancies between methods was due to the warping technique itself or due to the fact that the stimulus set sizes of studies using transformed images were unusually small.

Experiment 1 was the first experiment in which variance in stimulus facial masculinity came from both natural and artificial sources. The results of Experiment 1 showed that the way in which variance in masculinity is generated determines the direction of the relationship of masculinity and attractiveness. Image warping designed to cause changes in faces' perceived masculinity (artificial masculinity) in perceived masculinity, is negatively related to attractiveness, whereas the variation in perceived masculinity of untransformed men's faces (natural masculinity) is positively related to attractiveness. These results suggest that we should be careful in interpreting the results of studies in which images of faces are morphed. In particular, if the variation across images caused by morphing is assumed to represent a natural process, in this case facial growth, researchers should be careful to determine the validity of the assumption. In this case masculinization was not an ecologically valid simulation of natural variation in masculinity of men’s faces.

Study 2 was designed to explain the paradoxical finding of preference for masculinity and for feminization. The results of Study 2 showed that the way that people form judgments of masculinity of unaltered images of men is different from how they form judgments of masculinity of transformed images of men. Men’s facial masculinity is, to a small degree, a subset or reflection of the differences between men’s and women’s faces. By contrast, the variation in warped men’s faces that corresponds to differences in perceived masculinity does not parallel the differences between men’s and women’s faces. It is reasonable to conclude from the results of studies 1 and 2 that the masculinization model explains neither masculinity nor attractiveness judgments in terms of masculinity.

There are probably many reasons that the morphed facial images do not represent men’s facial masculinity. First, a morphed transition does not describe the changes in boys’ facial structure during puberty. Morphing is a smooth transformation between two forms. Growth is not a smooth transformation from child to adult. Different aspects of the body grow at different rates, for example the “growth curve” of the head is advanced relative to that of the face or that of the shoulders (Tanner, 1990). There is also individual variation in relative growth of different body parts, for example, only a fifth of boys reach their peak height velocity after genital development is complete (Cederquist, 1990). Growth is not predetermined and synchronous, it is stochastic and nonlinear.

Second, studies of how people differentiate men and women’s faces provides us with knowledge that contradicts the assumptions of the morphing model. Humans learn to differentiate men and women’s faces very accurately (Cheng, O’Toole, & Abdi, 2001; Golomb, Lawrence, & Sejnowski, 1991). Computational modeling of this ability indicates that to achieve reasonable accuracy the models must use many different dimensions of facial variation (Cheng, O'Toole, & Abdi, 2001). In Experiment 2 showed that 4 PLS components were needed to get good classification performance from the model. Cheng, O'Toole, & Abdi found that at least 7 principal components are necessary before performance increases asymptote. Thus, the factors that differentiate men and women are not described by a single dimension, as the morphing model assumes.

Averageness

Study 3 supported the averageness hypothesis with results from two different computational models of face recognition, PCA and ICA. PCA extracts sources of variation from image sets, which correspond to the configural differences among faces. ICA is similar to PCA, but it can identify correlated axes of variation in the image set, which include small image features, such as lines and curves, in addition to configural features that span the image frame, as does PCA.

Simulation results using either PCA or ICA showed that the similarity of a face image to an averaged face predicts its rated attractiveness. Effect sizes were similar for men and women. For both men and women similarity to a sex-congruent averaged face predicted their attractiveness whereas the similarity to the opposite-sex average did not, or did so to a lesser extent. The results also unexpectedly indicated that averageness – as an explanation for attractiveness – need not encode sexual dimorphism.

Langlois & Roggman (1990) demonstrated that increasing the numbers of faces in a pixel average increases the attractiveness of the resulting average, whether the average is of men's faces or women's faces. That Langlois & Roggman showed similar effects using men's and women's faces separately suggests that the averageness hypothesis is implicitly sexually dimorphic. Other researchers think it is likely that people construct subcategorical facial prototypes (Benson, 1995; Edelman, 1998b), so it is reasonable to suspect that averageness is sexually dimorphic. Study 3 showed that there may be no need to posit that individuals compare men's faces to a male average and women's faces to a female average. Study 3 showed instead that the combined male and female average accounted for as much variance in attractiveness as did sex-congruent average faces. In this way, averageness and sexual dimorphism are independent hypotheses of facial attractiveness.

Study 3 also used partial least squares regression (PLS) to determine 1) how facial attractiveness is reflected in the face image, and 2) whether human attractiveness judgments are replicable. PLS components do not necessarily correlate strongly with attractiveness ratings, but when combined linearly they generalize remarkably well in some training conditions. When combined nonlinearly the model generalizes almost perfectly.

The PLS modeling results strongly suggest that a very large part of attractiveness judgments do not depend on stimulus structure that differs for men and women. The model results suggest a two-factor model of attractiveness. The first factor describes both men and women's attractiveness. It is (almost) sufficient to account for women's attractiveness but not men's. The second factor is necessary to account for the remainder of men's attractiveness. Additionally, it describes both men's and women's facial attractiveness, despite making opposite predictions for men and women'. It is tempting to call the first factor averageness and the second sexual dimorphism, though further work will need to be done to determine whether this hypothesis is correct.

METHODOLOGICAL IMPROVEMENTS

The studies are informative for how they build upon existing methods, analytically and theoretically. Prior to this work, other researchers tested facial attractiveness hypotheses by treating faces as objects to be measured – thereby reducing psychological judgments to facial structure. Of course, faces are not simple objects; we do not experience faces like an anthropologist measures a skull. But facial recognition and judgments, such as masculinity, are measurements. Each involves a comparison of an judged stimuli with some idealized notion, whether it is a collection of previously seen people, or a concept such as masculinity. Our formal, careful, measurements of faces must therefore be more like our automatic perceptual measurements.

The divergence between the methods employed in this dissertation and the methods that have been used before reflects several differences in perspective and theoretical disposition. First, faces are complex objects (even when they don't move). Appropriate measurement methods must be able to handle large amounts of features as well as represent features in configurations. Rather than state a priori what a feature is, the computational methods used in this dissertation – ICA and PCA – use unsupervised feature extraction. PLS, on the other hand, uses guidance to extract features, but the guidance is not explicit; the researcher never identifies features except to align the images and define the cropping window. The PLS components are complex topographies that could not be measured manually.

Previous descriptions of faces were not consistent with our knowledge of how people perceive faces. Studies of facial recognition show that individuals cannot rely on feature distances to recognize faces (Burton, Bruce, & Dench, 1993; Davies, Ellis, & Shepherd, 1978). Furthermore, humans must use texture and surface (Bruce & Langton, 1994; Russell, 2004). We also know from face inversion studies that humans must use perceive faces as configurations of features (Thompson, 1980). Facial inversion does not change the distance between features but our perception of the face changes dramatically.

Computational models, such as PCA and ICA, are more consistent with what we know of face perception than are feature distance methods. This methodological distinction reflects that face perception and face attractiveness are two fields with little overlap. One purpose of this dissertation has been to show that these fields should complement rather than ignore each other.

A more fundamental difference is the conception of attractiveness. A common view of attractiveness is that it is a property of a face or person. In this view measurement is a fairly simple-minded exercise in extraction, and we assume that the aim of people's attractiveness judgment module is to accomplish this extraction. An alternate view is that attractiveness ratings are not properties of faces, but they are instead generated by perceivers of faces during their interaction with the stimulus. Moreover, the attractiveness judgments may be influenced by process factors of facial recognition, such as how quickly a stimulus is processed as a face. This is suggested as a general rule in preference formation, as Winkielman and Cacioppo have noted: "judgments reflect not only the descriptive factors, or what comes to mind, but also how things come to mind." (2001, p. 989).

EVOLUTIONARY EXPLANATIONS OF FACIAL ATTRACTIVENESS

In the introduction I outlined the model of facial attractiveness proposed by evolutionary psychologists (for example, Thornhill & Møller, 1997). There are reasons to abandon the immunological-hormonal model of attractiveness, but most of these are simply related to the fact that many of the assumptions of the immunological-hormonal model are contradicted by published research in endocrinology (see Chapter 1).

This is no reason, however, to assume that facial attractiveness has no evolutionary significance; it is difficult to imagine this could be true. It is not the purpose of this dissertation to argue against the correctness or relevance of ultimate (evolutionary) causation. As Langlois et al. (2000) reasoned, different explanations of attractiveness should be viewed as complementary rather than competitive.

The data in these studies are consistent with an alternate evolutionary account of attraction; attractiveness is partially a reflection of a flexible species recognition mechanism, based on general-purpose categorization mechanisms.

The importance of discriminating members of one's own species (conspecifics) from those of another species is vital. According to Ryan, Phelps, and Rand (2001), being able to select mates of one's own species is "the most crucial recognition task facing any sexually reproducing animal…" (p. 144). Whereas the consequence of failing to mate with a conspecific is serious, it is also a nontrivial problem to actually identify a conspecific. Vision researchers have long realized the difficulty in solving apparently simple problems of perception and recognition (Marr, 1982). The perceptual system that recognizes individuals should also be useful to discriminate species from non-species, so using a face recognition system as a species-recognition mechanism may often provide an economical and robust solution.

As shown in the results of Study 3, attractiveness judgments made to men's and women's faces seem to rest upon a similar logic. For example, PCA and ICA analyses showed that the degree of correspondence to a combined male+female averaged face predicts both men's and women's attractiveness. This suggests that attractiveness and facedness, or "humanness" are similar to some degree. A second architecture using partial least squares and a linear perceptron showed that a very large proportion of the variance in men's and women's attractiveness can be explained using the same logic for both men and women.

Although most faces we see every day are human, these stimuli vary in their degree of similarity to a prototypical human face. Faces that are similar to averaged face images are more attractive, which suggests that ease of processing contributes to their positive evaluation (Winkielman & Cacioppo, 2001). Indeed, in a study in which participants were asked to discriminate images of faces from images of scrambled face parts, they identified averaged faces most quickly, followed by attractive faces, and identification of unattractive faces was slowest (Rosen et al., in preparation). Valentine & Bruce (1986b) found that faces rated as "typical" were more quickly to be classified as faces than "atypical" faces. If faces are classified as unattractive by a species recognition mechanism, it could be characterized as an overgeneralization effect (Zebrowitz, 2003) if the judgment leads to falsely rejecting an otherwise suitable mate (that is, on the basis that the stimulus represents an individual of another species). Such a decision might have been adaptive in some environments, such as during the massive radiation of hominid species many millennia ago (Leakey, 1994), but not necessarily in our current context.

A species recognition system does not need to have innate representations of conspecifics (Lorenz's ducks did not have innate knowledge of conspecifics; they chose to identify Lorenz as their "mother" based on his early appearance in their lives). In fact it could be a bad strategy - evolutionarily - to have detailed innate knowledge of conspecifics. There is no evidence that human face recognition or attractiveness judgments are innate. For example, Gauthier and Nelson (2001) noted that researchers have found infant preferences for facelike stimuli as newborns and at 2 months of age, but that researchers have not found face preferences for infants between these ages. The evidence on infant attractiveness preferences is even less supportive of an innate view. Although young infants (6mo) show evidence that they prefer attractive faces to unattractive faces (Langlois et al., 1997), newborns do not (Kalakanis, 1997).

Moreover, although adults' attractiveness judgments are assumed to be cross-culturally invariant (Fink, Grammer, & Thornhill, 2001), this is not true. Although there is generally cross-cultural agreement, it is not as high as within culture agreement. Langlois et al. (2000) analyzed the results of many cross-cultural studies metaanalytically; the average correlation between raters from different cultures was .54. The effective cross-cultural reliability was higher, $r = .88$, but effective reliabilities should not interpreted as correlation coefficients (Rosenthal, 1991).

The "other race" effect can account for cultural variation in attractiveness standards. People adapt to local variation in facial appearance so that they may distinguish individuals, but expertise with local variation does not translate into proficiency with distinguishing individuals of an unfamiliar race; individuals of an unfamiliar race are perceived as looking very similar to each other. Perceptual learning accounts for such effects; it "involves differentiation of distinctive features" (Gibson, 1969, p.146) that are pieced together through experience with a class of stimuli. When individuals are presented with pictures of people of an unfamiliar race, the distinguishing features change and recognition rates are lowered (O'Toole, Deffenbacher, Valentin, & Abdi, 1994). With training, individuals can learn to distinguish individuals of a previously unfamiliar race (Goldstein & Chance, 1985), so it is assumed that they have learned the unique features that distinguish individuals in the newly-familiarized race.

Malinowski, a student of Wundt and pioneer of the cultural mastery style of anthropology in which the ethnographer must become immersed in the culture he or she observes, described his personal acculturation during field work among the Trobriand Islanders of Melanisia. Some of his writings suggest a direct relationship between the other-race effect and cultural standards of beauty:

> "…I was less susceptible at first to individual differences and more impressed by the general type. But with greater familiarity, I came to feel that too dark or too yellow a skin, to straight or too frizzy hair, a mouth as thin as that of a European, and an aquiline nose were features unpleasant in a Melanesian. At the same time I became able to appreciate beauty within the racial type and de facto always knew more or less who would be attractive to a native, and who not." (Malinowski, 1929, p. 308)

While Malinowski became accustomed to the appearance of the native people, the Trobrianders did not appear to have adapted to the appearance of Europeans during Malinowski's visit:

> "Europeans, the natives frankly say, are not good looking. …they were quick to add that the ethnographer was a meritorious exception. …they always told me that I looked much more like a Melanisian than like an ordinary white man." (Malinowski, 1929, p. 307).

The other-race effect may be a reflection of attractiveness as a species-recognition mechanism. If true, it could partially explain why so many ethnographic accounts show individuals of many cultures dehumanize individuals of other cultures or races. It has been suggested that ethnic groups "essentialize" humans, perhaps even considering outgroup members to be of different species (Gil-White, 2001). Dehumanization often co-occurs with intercultural conflict. Schultze (1907) described the prejudices of the Hottentots of Africa, who hypothesized that a nearby tribal group was descended from baboons. The Nazis in 1930s Germany vilified German Jews, by portraying them as corrupt "vermin." Nazi propagandists caricatured Jewish facial features, producing nightmarish images of subhuman creatures. The pernicious objective was for ordinary Germans to dehumanize their Jewish countrymen and neighbors (Goldhagen, 1996).

Actual, acute appearance differences can also produce a dehumanization effect. For example, many pre-industrial cultures practice teeth alteration, such as using natural substances to permanently turn teeth black, chipping away at the teeth to make them sharp and triangular, or removing specific teeth altogether. Their facial appearance - and thus, their facial prototypes - can be strikingly different such that those who are unfamiliar with such alterations may react with revulsion. Conversely, to people who practice teeth alteration, unaltered white teeth are repulsive. Among the Tiv of Africa, "…a woman would refuse you for not having your teeth cut, and rail at you for having flat teeth like a monkey or a foreigner…" (Akiga, 1939, p. 47). In Oceania, people who blacken their teeth call individuals with white teeth "dogs" (for example, Adriani, & Kruyt, 1951; Wilken, 1893; Kennedy, 1942). People within these cultures are so accustomed to the appearance of individuals with altered teeth that, to them, unaltered teeth are a mark of animals or young children, not adult humans or potential mates.

This evolutionary interpretation of averageness theory of facial attractiveness as a species recognition mechanism differs from traditional evolutionary psychology theories. First, it does not assume that the neural systems underlying humans' attractiveness judgments is "hardwired" or "innate." It may function more efficiently if it is a self-organizing system that accumulates knowledge about faces through perceptual learning. Second, it does not assume that the variance in attractiveness judgments among conspecifics is related to genetic quality – at least the portion of attractiveness judgments attributable to averageness. Third, the system responsible for attractiveness judgments is not modularized – components of the same system for distinguishing one friend from another are partially responsible for generating initial attractiveness judgments. Moreover, the system is simply specialized recognition "hardware," which can be adapted to make different perceptual distinctions.

CONCLUSION

The purpose of this dissertation was to provide a critique of the explanatory framework of contemporary facial attractiveness hypotheses and to show that there is a viable alternative. In Chapter 1 I showed that the theoretical framework, although internally consistent, is a complex system whose components are not supported by research in relevant fields (see, for example, Figure 4). I then focused on a particular aspect of the theoretical structure, the assumptions the immuno-endocrinological model about hormonal influences on facial growth, showing that the endocrinology literature does not support such assumptions.

Surprisingly, the literature on facial perception seems to be not well-integrated with facial attractiveness research; much of our knowledge about face recognition is relevant to a proper understanding of facial attractiveness. For example, the methods used in face perception research could be useful to test facial attractiveness theories. Essentially, models of cognitive representation can double as facial measurement methods, improving facial attractiveness research methods.

I presented three studies, each of which was designed to test the averageness and sexual dimorphism hypotheses of facial attractiveness. Additionally, each study was designed as an example of how facial recognition methods can be used to test facial attractiveness theories. Results of the three studies showed first, averageness and sexual dimorphism are independent explanations for why faces are attractive. I also demonstrated that the way in which variance in men's facial masculinity is generated determines the direction of the relationship between masculinity and attractiveness. Second, it is likely that averageness should be measured as the degree of similarity between a face and a hypothetical gender-neutral prototype, rather than a sub-categorical prototype such as "averaged man." Third, unsupervised learning algorithms (principal components analysis and independent components analysis) can explain moderate amounts of variance in attractiveness, and supervised learning architectures can explain all of the variance, creating a mechanism that mimics human judgment of attractiveness.

The important next steps in the study of physical attractiveness are first, investigating the time-course dynamics of attraction. Perceptual fluency with individual faces may be useful to help explain variation in interpersonal attraction between two people over time. Second, individual differences in judgments of attractiveness should be investigated. As attractiveness is assumed to have certain - almost economic – value in the "dating marketplace," (Berscheid & Walster, 1974) and, as shown here, it is possible to create a system that mimics attractiveness judgments, we should seek to understand why it is that people do not agree with each other more strongly.

It is known that experience has canalizing effects on perception, such as explored in studies of plasticity (Hubel, Wiesel, & LeVay, 1977; Miikkulainen, Bednar, Choe, & Sirosh, 1997) and the other-race effect on facial recognition (for example, Goldstein & Chance, 1985). It is reasonable to hypothesize that the somewhat peculiar assortment of faces an individual experiences could account for differences in perception of the correspondence of stimuli to a facial prototype, differences in perception of masculinity/femininity, and therefore differences in perceptions of attractiveness.

Appendix

Elliptical window cropping was performed by adjusting the formula for an ellipse. The horizontal and vertical boundaries of faces within images were measured (four points per face) using NIH Image.

To find the horizontal and vertical radii and vertical elevation for an ellipse that would fit the entire set of faces, the pixel coordinates from NIH Image were transformed into a zero-centered coordinate system. For each of the four point locations, a variation on the formula for an ellipse was used to indicate the deviation between face shape and ellipse shape.

The squared deviation between ellipse boundary and top of forehead, where α and β are the vertical and horizontal ellipse radii, δ is the vertical displacement of the ellipse, $x_1 y_1$ is the forehead coordinate:

$$\left(-\delta+\sqrt{1-\frac{x_1^2}{\alpha^2}\beta^2}-y_1\right)^2$$

The squared horizontal deviation between an ellipse and the right side of a face:

$$\left(-1\sqrt{1-\frac{y_2^2}{\beta^2}\alpha^2}-x_2\right)^2$$

The squared horizontal deviation between an ellipse and the left side of a face:

$$\left(-1\sqrt{1-\frac{y_3^2}{\beta^2}\alpha^2}-x_3\right)^2$$

The squared vertical deviation between an ellipse and the bottom of a face:

$$\left(\delta+\sqrt{1-\frac{x_4^2}{\alpha^2}\beta^2}-y_4\right)^2$$

Excel's solver was used to find values of α, β, and δ that minimized the summed deviations between facial measurements of the set of images and the location of the ellipse.

For computational modeling, only pixels within the elliptical window were used in analysis. A series of PCA analyses were undertaken to determine how choosing windows of arbitrary size and shape would affect predictions of attractiveness, and whether the best predictions of attractiveness also were in analyses in which optimal window shapes were chosen.

X_M (averaged male face)		Vertical						
		40	50	60	70	80	90	100
Horizontal	40	.30	.25	.23	.20	.18	.20	.18
	50	.33	.34	.32	.26	.25	.29	.24
	60	.35	.36	.34	.27	.27	.34	.24
	70	.38	.37	.35	.28	.27	.31	.25
	80	.45	.44	.42	.37	.36	.38	.38
	90	.46	.42	.45	.44	.40	.43	.45
	100	.44	.45	.47	.44	.39	.43	.43
X_{MF} (averaged male+female face)		Vertical						
		40	50	60	70	80	90	100
Horizontal	40	.25	.21	.26	.21	.18	.20	.18
	50	.23	.25	.30	.25	.24	.27	.26
	60	.21	.23	.33	.26	.25	.27	.28
	70	.31	.36	.40	.29	.26	.24	.25
	80	.35	.32	.35	.27	.23	.25	.27
	90	.36	.30	.34	.30	.27	.32	.32
	100	.35	.29	.33	.29	.28	.32	.31

Table 5: Predicting men's attractiveness under varied ellipse parameters.

Optimal ellipse parameters for men were $\alpha = 102$, $\beta = 81$, and $\delta = -3$. For prediction of men's attractiveness, adjusting the horizontal parameter covaried more strongly with variation in prediction of attractiveness. Prediction of men's attractiveness did not correlate with degree of ellipse fit, $r = .01$, *ns*.

X_F (averaged female face)

		Vertical						
		40	50	60	70	80	90	100
Horizontal	40	.29	.26	.31	.37	.36	.37	.35
	50	.35	.32	.37	.42	.42	.40	.38
	60	.38	.35	.43	.44	.44	.39	.34
	70	.37	.39	.42	.41	.40	.33	.32
	80	.31	.32	.36	.38	.34	.32	.30
	90	.27	.28	.32	.36	.38	.36	.34
	100	.28	.29	.32	.34	.36	.36	.35

X_{MF} (averaged male+female face)

		Vertical						
		40	50	60	70	80	90	100
Horizontal	40	.32	.31	.32	.34	.34	.33	.38
	50	.31	.32	.35	.36	.35	.36	.40
	60	.34	.33	.37	.38	.40	.41	.40
	70	.33	.40	.43	.48	.47	.46	.41
	80	.37	.42	.46	.50	.45	.42	.38
	90	.36	.41	.47	.49	.46	.42	.39
	100	.35	.41	.43	.43	.42	.39	.36

Table 6: Predicting women's attractiveness under varied ellipse parameters.

Optimal ellipse parameters for women were $\alpha = 84$, $\beta = 74$, and $\delta = +10$. Prediction of women's attractiveness correlated with degree of ellipse fit, $r = .71$, $p < .001$.

For other computational analyses in the dissertation, one elliptical window was used that minimized the sum of squared pixel deviations for unaltered images of both men and women: $\alpha = 93$, $\beta = 74$, and $\delta = 8$.

References

Akiga (1939). Akiga's Story: The Tiv Tribe as Seen by one of its Members. London: Oxford University Press.

Adriani, N. & Kruyt, A.C. (1951). The Bare'e-Speaking Toradja of Central Celebes. Amsterdam: Noord-Hallandsche Uitgevers Maatschappij.

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

Barlow, H.B. (1989). Unsupervised learning. *Neural Computation, 1,* 295-311.

Bartlett, M.S. (2001). *Face Image Analysis by Unsupervised Learning.* Boston, MA: Kluwer Academic Publishers.

Benson, P.J. (1995). Perspectives on face perception: Directing research by exploiting emergent prototypes. In T. Valentine (Ed.) *Cognitive and Computational Aspects of Face Recognition: Explorations in Face Space.* (pp. 204-224). New York, NY: Routledge.

Berscheid, E., & Walster, E. (1974). Physical Attractiveness. In L. Berkowitz (Ed.) *Advances in experimental social psychology*. (pp. 157-215). New York: Academic Press.

Bell, A.J., & Sejnowski, T.J. (1997). The "independent components" of natural scenes are edge filters. *Vision Research, 37,* pp. 3327-3338.

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

Bronstad, P.M., Ramsey, J.L., & Langlois, J.H. (2002, May). *Sample size explains discrepancies in facial attractiveness research: Masculine male faces are more attractive.* Paper presented at the American Psychological Society, New Orleans, LA.

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

Bruce, V., Burton, A.M., & Hancock, P. (1995). Missing dimensions of distinctiveness. In T. Valentine (Ed.) *Cognitive and computational aspects of face recognition: Explorations in face space.* (pp. 138-158). New York: Routledge.

Bruce, V., & Langton, S. (1994). The use of pigmentation and shading information in recognising the sex and identities of faces. *Perception, 23,* 803-822.

Bruce, V., & Young, A. (1998). *In the eye of the beholder: The science of face perception*. New York: Oxford University Press.

Burton, A.M., Bruce, V., & Dench, N. (1993). What's the difference between men and women? Evidence from facial measurement. *Perception, 22,* 153-176.

Burton, A.M., Miller, P., Bruce, V., Hancock, P.J.B., & Henderson, Z. (2001). Human and automatic face recognition: A comparison across image formats. *Vision Research, 41,* 3185-3195.

Busey, T.A. (2001). Formal models of familiarity and memorability in face recognition. In M.J. Wenger, & J.T. Townsend (Eds.) *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges*. (pp. 147-191). Mahwah, NJ: Lawrence Erlbaum.

Buss, D.M; Schmitt, D.P. (1993). Sexual strategies theory: An evolutionary perspective on human mating. *Psychological Review, 100,* 204-232.

Cassia, V.M., Turati, C., & Simion, F. (2004). Can a nonspecific bias toward top-heavy patterns explain newborns' face preference? *Psychological Science, 15,* 379-383.

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., & Cohen, P. (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Eflbaum.

Cottrell, G.W., Dailey, M. N., Padgett, C., Adolphs, 2001, R. (2001). Is all face processing holistic? The view from UCSD. In M.J. Wenger, & J.T. Townsend (Eds.) *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges*. (pp. 347–395). Mahwah, NJ: Lawrence Erlbaum.

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,* 925-935.

Cunningham, M.R., Barbee, A.P., & Philhower, C.L. (2002). Dimensions of facial physical attractiveness: The intersection of biology and culture. In G. Rhodes and L.A. Zebrowitz (Eds.) *Facial Attractiveness: Evolutionary, Cognitive, and Social Perspectives.* (pp. 193-238). Westport, CN: Ablex.

Cunningham, M.R., Barbee, A.P., & Pike, C.L. (1990). What do women want? Facialmetric assessment of multiple motives in the perception of male facial physical attractiveness. *Journal of Personality and Social Psychology, 59,* 61-72.

Cunningham, M.R., Druen, P.B., & Barbee, A.P. (1997). Angels, mentors, and friends: Trade-offs among evolutionary, social, and individual variables in physical appearance. In J. A. Simpson and D. T. Kenrick, Eds., *Evolutionary Social Psychology.* (pp. 109-140). Mahwah, NJ: Erlbaum.

Cunningham, M.R., Roberts, A.R., Barbee, A.P., Druen, P.B., & Wu, C. (1995). "Their ideas of beauty are, on the whole, the same as ours": Consistency and variability in the cross-cultural perception of female physical attractiveness. *Journal of Personality and Social Psychology, 68,* 261 – 279.

Davies, G.M., Ellis, H.D., & Shepherd, J.W. (1978). Face recognition accuracy as a function of mode of representation. *Journal of Applied Psychology, 63,* 180-187.

Dhabhar, F.S. (1998). Stress-induced enhancement of cell-mediated immunity. *Annals of the New York Academy of Sciences, 840,* 359-372.

Edelman, S. (1998a). Representation is representation of similarities. *Behavioral and Brain Sciences, 21,* 449-498.

Edelman, S. (1998b). Spanning the face space. *Journal of Biological Systems, 6,* 265-280.

Efron, B. & Tibshirani, R.J.. 1993 *An Introduction to the Bootstrap.* Chapman & Hall: New York.

Enlow, D.H. (1990). Control processes in facial growth. In *Facial growth, 3rd Edition*. Donald H. Enlow (Ed.) (pp. 229-248). Harcourt Brace Jovanovich, Inc. Philadelphia.

Enquist, M., Ghirlanda, S., Lundqvist, D., & Wachtmeister, C., (2002). An ethological theory of attractiveness. In G. Rhodes and L.A. Zebrowitz (Eds.) *Facial attractiveness: Evolutionary, cognitive, and social perspectives*. (pp. 127-151). Westport, CN: Ablex.

Etcoff, N. (1999). *Survival of the prettiest.* New York: Doubleday.

Farkas, L.G. (1981). *Anthropometry of the head and face in medicine.* New York, New York: Elsevier North Holland.

Fink, B., Grammer, K., & Thornhill, R. (2001). Human (*Homo sapiens*) facial attractiveness in relation to skin texture and color. *Journal of Comparative Psychology, 115,* 92-99.

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

Galton, F. (1879). Composite portraits. *Journal of the Anthropological Institute of Great Britain and Ireland, 8,* 132-142.

Gangestad, S.W. & Thornhill, R. (1997). Human sexual selection and developmental stability. In J. Simpson & Kenrick (Eds.), *Evolutionary social psychology,* 169—195.

Gangestad, S. W., Thornhill, R., & Garver, C. E. (2002). Changes in women's sexual interests and their partners' mate retention tactics across the menstrual cycle: Evidence for shifting conflicts of interest. *Proceedings of the Royal Society of London, Series B, 269,* 975-982.

Gauthier, I. & Nelson, C.A. (2001). The development of face expertise. *Current Opinion in Neurobiology, 11,* 219-224.

Geladi, P., & Kowalski, B.R. (1986). Partial least-squares regression: A tutorial. *Analytica Chimica Acta, 185,* 1-17.

Gibson, E.J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.

Gil-White, F.J. (2001). Are ethnic groups biological "species" to the human brain? Essentialism in our cognition of some social categories. *Current Anthropology, 42,* 515-554.

Goldhagen, D.J. (1996). *Hitler's Willing Executioners: Ordinary Germans and the Holocaust.* New York, NY: Alfred A. Knopf.

Goldstein, A.G., & Chance, J.E. (1985). Effects of training on Japanese face recognition: Reduction of the other-race effect. *Bulletin of the Psychonomic Society, 23,* 211-214.

Golomb, B., Lawrence, D., & Sejnowski, T. (1991). Sexnet: A neural network identifies sex from human faces. In Lippman, R., Moody, J., and Touretzky, D. (Eds.) *Advances in Neural Information Processing Systems, Volume 3*. (pp. 572-577). Morgan-Kaufmann, San Mateo, CA.

Grammer, K., Fink, B., Juette, A., Ronzal, G., & Thornhill, R. (2002). Female faces and bodies: N-dimensional feature space and attractiveness. In G. Rhodes and L.A. Zebrowitz (Eds.) *Facial attractiveness: Evolutionary, cognitive, and social perspectives.* (pp. 91-125). Westport, CN: Ablex.

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

Grumbach, M.M. (2000). Pubertal maturation in aromatase deficiency and resistance to estrogen. In J.P. Bourguignon and T.M. Plant (Eds.), *The onset of puberty in perspective.* (pp. 247 – 267). New York: Elsevier.

Grumbach, M.M. & Auchus, R.J. (1999). Estrogen: Consequences and implications of human mutations in synthesis and action. *Journal of Clinical Endocrinology and Metabolism, 84,* 4677–4694.

Hagan, M.T., Demuth, H.B., & Beale, M. (1996). *Neural Network Design*. PWS Publishing Co, Boston, MA.

Hancock, P.J.B., Bruce, V., & Burton, M.A. (1998). A comparison of two computer-based face identification systems with human perceptions of faces. *Vision Research, 38,* 2277-2288.

Hasselquist, D., Marsh, J.A., Sherman, P.W., & Wingfield, J.C. (1999). Is avian humoral immunocompetence suppressed by testosterone? *Behavioral Ecology and Sociobiology, 45,* 167-175.

Hillgarth, N. & Wingfield, J. (1997). Parasite-mediated sexual selection: Endocrine aspects. In D. Clayton and J. Moore (Eds.) *Host-Parasite Evolution: General Principles and Avian Models* (pp. 78-104).Oxford: Oxford University Press.

Huang, X., & Pan, W. (2003). Linear regression and two-class classification with gene expression data. *Bioinformatics, 19,* 2072-2078.

Hubel, D.H., Wiesel, T.N., & LeVay, S. (1977). Plasticity of ocular dominance columns in monkey striate cortex. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 278,* 377-409.

Hume, D. (1985). *Essays: Moral, Political, and Literary* (E. Miller, Ed.) Indianapolis, IN: LibertyClassics. (Original work published 1777).

Hunter, W.S., Balbach, D.R., & Lamphiear, D.E. (1970). The heritability of attained growth in the human face. *American Journal of Orthodontics, 58*, 128-134.

Iliffe, A.H. (1960). A study of preferences in feminine beauty. *British Journal of Psychology, 51*, 267-273.

Itin, P.H., Kirtschig, G., Gilli, L., & Happle, R., (1997). Superciliary upsweep or tented eyebrows. *Journal of the American Academy of Dermatology, 37,* 295-297.

Johnston, V.A., Hagel, R., Franklin, M., Fink, B., Grammer, K. (2001). Male facial attractiveness: Evidence for hormone-mediated adaptive design. *Evolution and Human Behavior, 22,* 251-267.

Johnston, R.A., & Ellis, H.D. (1995). The development of face recognition. In T. Valentine (Ed.), *Cognitive and computational aspects of face recognition: Explorations in face space.* (pp. 1-23). London: Routledge.

Jones, D. (1996). *Physical attractiveness and the theory of sexual selection: Results from five populations*. Ann Arbor, MI: Museum of Anthropology, University of Michigan.

Kalakanis, L.E. (1997). Newborns' preferences for attractive faces. (Doctoral Dissertation, The University of Texas at Austin, 1997). *Digital Dissertations, AAT 9824985*.

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

Kennedy, R. (1942). *The Ageless Indies.* New York: The John Day Company.

Kirkpatrick, M. (1996). Good genes and direct selection in the evolution of mating preferences. *Evolution, 50*, 2125-2140.

Langlois, J.H., Roggman, L.A., Casey, R.J., Ritter, J.M., Rieser-Danner, L.A., & Jenkins, V.Y. (1987). Infant preferences for attractive faces: Rudiments of a stereotype? *Developmental Psychology, 23,* 363-369.

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

Langlois, J.H., Roggman, L.R., & Musselman, L. (1994). What's average and not average about attractive faces? *Psychological Science, 5,* 214-220.

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

Leakey, R.E. (1994). *The Origin of Humankind.* New York: Basic Books.

Lee, K., Byatt, G., & Rhodes, G. (2000). Caricature effects, distinctiveness, and identification: Testing the face-space framework. *Psychological Science, 11,* 379-385.

Leopold, D., O'Toole, A.J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience, 4,* 89-94.

Lindzey, G. (1965). Morphology and behavior. In G. Lindzey and C.S. Hall (Eds.) *Theories of Personality: Primary Sources and Research* (pp. 344-353). New York: Wiley and Sons, Inc..

Little, A.C., Penton-Voak, I.S., Burt, D.M., & Perrett, D.I. (2002). Evolution and individual differences in the perception of attractiveness: How cyclic hormonal changes and self-perceived attractiveness influence female preference for male faces. In G. Rhodes and L.A. Zebrowitz (Eds.) *Facial attractiveness: Evolutionary, cognitive, and social perspectives.* (pp. 59-90). Westport, CN: Ablex.

MacLin, O.H. & Webster, M.A. (2001). Influence of adaptation on the perception of distortions in natural images. *Journal of Electronic Imaging, 10,* 100-109.

Malinowski, B. (1929). *The Sexual Lives of Savages in North-Western Melanesia.* New York: Horace Liveright.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* San Francisco, CA: W.H. Freeman.

McArthur, L.Z., & Berry, D.S. 1987 Cross-cultural agreement in perceptions of babyfaced adults. *Journal of Cross-cultural Psychology,* 18*,* 165-192.

McKusick, V.A. (1992). *Mendelian inheritance in man: Catalogs of autosomal dominant, autosomal recessive, and X-linked phenotypes*. Baltimore, MD: The Johns Hopkins University Press.

Meyer, D.A., & Quong, M.W. (1999). The bio/logic of facial geometry. *Nature, 397,* 661-662.

Miikkulainen, R., Bednar, J.A., Choe, Y., & Sirosh, J. (1998). Self-organization, plasticity, and low-level visual phenomena in a laterally-connected map model of the primary visual cortex. In R.L. Goldstone, P.G. Schyns, and D.L. Medin (Eds.) *Psychology of Learning and Motivation, volume 36:Perceptual Learning,* (pp. 257-308). San Diego, CA: Academic Press.

Moon, H. & Phillips, P.J. (2001). Computational and performance aspects of PCA-based face recognition algorithms. *Perception, 30,* 301-321.

O'Toole, A.J., Abdi, H., Deffenbacher, K.A., & Valentin, D., (1995). A perceptual learning theory of the information in faces. In T. Valentine (Ed.). *Cognitive and computational aspects of face recognition: Explorations in face space.* (pp. 159-182). New York: Routledge.

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

O'Toole, A. J., Deffenbacher, K. A, Valentin, D., & Abdi, H.(1994). Structural aspects of face recognition and the other-race effect. *Memory & Cognition, 22,* 208-224.

O'Toole, A.J., Price, T., Vetter, T., Bartlett, J.C., & Blanz, V. (1998). 3D shape and 2D surface textures of human faces: The role of "averages" in attractiveness and age. *Image and Vision Computing Journal, 18,* 9-19.

O'Toole, A.J., Wenger, M.J., & Townsend, J.T. (2001). Quantitative models of perceiving and remembering faces: Precedents and possibilities. In M.J. Wenger, & J.T. Townsend (Eds.) *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges*. (pp. 1-38). Lawrence Erlbaum, Mahwah, NJ.

O'Toole, A.J., Vetter, T., Troje, N.F., & Bülthoff, H.H. (1997). Sex classification is better with three-dimensional head structure than with image intensity information. *Perception, 26,* 75-84.

Palmer, P.E.S., & Reeder, M.M. (2001). *The Imaging of Tropical Diseases, with Epidemiological, Pathological and Clinical Correlation*. Retrieved August 4, 2004, from http://tmcr.usuhs.mil/toc.htm.

Palmer, S.E. (1978). Fundamental aspects of cognitive representation. In E. Rosch and B.B. Lloyd (Eds.) *Cognition and categorization.* (pp. 259-303). Hillsdale, NJ: Lawrence Erlbaum Associates.

Pellicer & Burke (2002, February 11). Growing number of people are victims of health care crisis. *The Olympian*.

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

Penton-Voak, I.S. & Perrett, D.I. (2000). Consistency and individual differences in facial attractiveness judgments: An evolutionary perspective. *Social Research, 67,* 219-245.

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

Penton-Voak, I.S., Jones, B.C., Little, A.C., Baker, S., Tiddeman, B., Burt, D.M., & Perrett, D.I. (2001). Symmetry, sexual dimorphism in facial proportions and male facial attractiveness. *Proceedings of the Royal Society of London, Series B, 268,* 1617-1623.

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

Perrett, D.I., & Penton-Voak, I. (1999). The biologic of facial geometry (reply to Meyer & Quong). *Nature, 397,* 662.

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

Pollard, J., Shepherd, J., Shepherd, J. (1999). Average faces are average faces. *Current Psychology: Developmental, Learning, Personality, Social, 18,* 98-103.

Reis, V.A. & Zaidel, D.W. (2001). Brain and face: Communicating signals of health in the left and right sides of the face. *Brain and Cognition, 46,* 240-244.

Rhodes, G., Yoshikawa, S., Clark, A., Lee, K., McKay, R., Akamatsu, S. (2001). Attractiveness of facial averageness and symmetry in non-Western cultures: In search of biologically based standards of beauty. *Perception, 30,* 611-625.

Rhodes, G., Hickford, C., & Jeffery, L. (2000). Sex-typicality and attractiveness: Are supermale and superfemale faces super-attractive? *British Journal of Psychology, 91,* 125-140.

Rhodes, G., Sumich, A., & Byatt, G., (1999). Are average facial configurations attractive only because of their symmetry? *Psychological Science, 10,* 52-58.

Rhodes, G., & Zebrowitz, L. A. (2002). Introduction. In G. Rhodes and L.A. Zebrowitz (Eds.) *Facial attractiveness: Evolutionary, cognitive, and social perspectives.* (pp. vii-x). Westport, CN: Ablex.

Rosenthal, R. (1991). *Meta-Analytic Procedures for Social Research*. Newbury Park, CA: Sage.

Rubenstein, A.R., Kalakanis, L.A., & Langlois, J.H. (1999). Infant preferences for attractive faces: A cognitive explanation. *Developmental Psychology, 35,* 848-855.

Rubenstein, A.J., Langlois, J.H., & Roggman, L.A. (2002). What makes a face attractive and why: The role of averageness in defining facial beauty. In G. Rhodes and L.A. Zebrowitz (Eds.) *Facial attractiveness: Evolutionary, cognitive, and social perspectives.* (pp. 1-33). Westport, CN: Ablex.

Russell, R. (2004, May). *Male and female faces are differently pigmented.* Paper presented at the American Psychological Society, Chicago, IL.

Ryan, M.J. (1997). Sexual selection. In J.R. Krebs and N.B. Davies (Eds.) *Behavioral Ecology: An Evolutionary Approach* (pp. 179-202). Malden, MA: Blackwell.

Ryan, M.J., Phelps, S.M., & Rand, A.S. (2001). How evolutionary history shapes recognition mechanisms. *TRENDS in Cognitive Sciences, 5,* 143-148.

Saunders, S.R., Popovich, F., & Thompson, G.W. (1980). A family study of craniofacial dimensions in the Burlington growth centre sample. *American Journal of Orthodontics, 78,* 394 – 403.

Scheib, J.E., Gangestad, S.W., & Thornhill, R. (1999). Facial attractiveness, symmetry, and cues of good genes. *Proceedings of the Royal Society of London, Series B, 266,* 1913-1917.

Schultze, L. (1907). *In Namaland and the Kalahari.* Jena: Gustav Fischer.

Shackelford, T.K., & Larsen, R.J. (1999). Facial attractiveness and physical health. *Evolution and Human Behavior, 20,* 71-76.

Shepherd, J. (1981). Social factors in face recognition. In G. Davies, H. Ellis, and J. Shepherd (Eds.) *Perceiving and Remembering Faces* (pp. 55-79). New York, NY: Academic Press.

Smith, J.D. (2002). Exemplar theory's predicted typicality gradient can be tested and disconfirmed. *Psychological Science, 13,* 437 – 442.

Stevenage, S. V. (1995). Expertise and the caricature advantage. In T. Valentine (Ed.) *Cognitive and computational aspects of face recognition: Explorations in face space*. (pp. 24-46). New York: Routledge.

Swaddle, J.P. & Ruff, D.A. (2004). Starlings have difficulty in detecting dot symmetry: Implications for studying fluctuating asymmetry. *Behaviour, 141,* 29-40.

Symons, D. (1995). Beauty is in the adaptations of the beholder: The evolutionary psychology of human female sexual attractiveness. In P.R. Abramson and S.D. Pinkerton (Eds.) *Sexual Nature/Sexual Culture*. University of Chicago Press: Chicago.

Tanner, J.M. (1990). *Foetus into man: Physical growth from conception to maturity.* (Rev. ed.). Cambridge, MA: Harvard University Press.

Thesleff, I. (1997). Homeobox genes and growth factors in regulation of craniofacial and tooth morphogenesis. *Acta Odontologica Scandinavica, 53,* 129 – 134.

Thompson, P. (1980). "Margaret Thatcher: A new illusion." *Perception, 9,* 483-484.

Thornhill, R., & Gangestad, S.W. (1993). Human facial beauty: Averagness, symmetry, and parasite resistance. *Human Nature, 4,* 237-269.

Thornhill, R., & Gangestad, S.W. (1999). Facial attractiveness. *Trends in Cognitive Sciences, 3,* 452 – 460.

Thornhill, R., & Grammer, K. (1999). The body and face of woman: One ornament that signals quality? *Evolution and Human Behavior, 20,* 105-120.

Thornhill, R., & Møller, A.P. (1997). Developmental stability, disease and medicine. *Biological Reviews, 72,* 497-548.

Tinbergen, N. (1976). *The Study of Instinct.* New York: Oxford University Press.

Tooby, J. & Cosmides, L. (1992). The psychological foundations of culture. In J.H. Barkow, L. Cosmides, and J. Tooby (Eds.) *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (pp. 19-136). New York, NY: Oxford University Press.

Turk, M.A. & Pentland, A.P. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience, 3,* 71-86.

Valentin, D., & Abdi, H. (1996). Can a linear autoassociator recognize faces from new orientations? *Journal of the Optical Society of America A, 13,* 717-724.

Valentin, D., Abdi, H., Edelman, B., & O'Toole, A.J. (1997). Principal component and neural network analyses of face images: What can be generalized in gender classification? *Journal of Mathematical Psychology, 41,* 398-413.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology, 79,* 471-491.

Valentine, T., Bruce, V. (1986a). Recognizing familiar faces: The role of distinctiveness and familiarity. *Canadian Journal of Psychology, 40,* 300-305.

Valentine, T., Bruce, V. (1986b). The effects of distinctiveness in recognising and classifying faces. *Perception, 15,* 525-535.

Valentine, T., Chiroro, P. & Dixon, R. (1995). An account of the own-race bias and the contact hypothesis based on a 'face space' model of face recognition. In T. Valentine (Ed.) *Cognitive and computational aspects of face recognition.* (pp. 1-23). London: Routledge.

Wedekind, C., & Füri, S. (1997). Body odour preferences in men and women: Do they aim for specific MHC combinations or simply heterozygosity? *Proceedings of the Royal Society of London, Series B, 264,* 1471-1479.

Wentz, L. (2002, Nov. 11). P&G and GM lead global ad spending. *Advertising Age, 73,* 28.

Wild, H.H., Barrett, S.E., Spence, M.J., O'Toole, A.J., Cheng, Y.D., & Brooke, J. (2000). Recognition and sex categorization of adults' and childrens' faces in the absence of sex stereotyped cues. *Journal of Experimental Child Psychology, 77*, 261-299.

Wilken, G.A. (1893). *Manual for the Comparative Ethnology of the Netherlands East Indies.* Leiden: E.J. Brill.

Winkielman, P. & Cacioppo, J.T. (2001). Mind at ease puts a smile on the face: Psychophysiological evidence that processing facilitation elicits positive affect. *Journal of Personality and Social Psychology, 81,* 989-1000.

Young, A.W., & Bruce, V. (1991). Perceptual categories and the computation of "grandmother." *European Journal of Cognitive Psychology, 3,* 5-49.

Zebrowitz, L.A. (2003). Overgeneralization effects in perceiving nonverbal behavior: Evolutionary and ecological origins. *Journal of Nonverbal Behavior, 27,* 133-138.

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

Vita

Philip Matthew Bronstad was born September 26, 1972 in Ann Arbor, Michigan, to Patricia Margie Bronstad and Gilbert Walton Bronstad. The family moved to Texas, New Zealand, and Oregon, before settling back in Austin, Texas in 1990. After graduating from Austin High School, he enrolled at the University of Texas at Austin in 1991, earning a Bachelor of Arts in May 1997 with special honors in psychology. In the fall of 1997 he began working on a Ph.D. in psychology at the University of Texas at Austin. He and Cheryl Annette Browne were married in April 2001.

Permanent address: 3305 Helms St.
Austin, TX 78705

This dissertation was typed by the author.