2011

The role of facial appearance in gender categorization

Joshua J. Simpkins
University of Central Florida

THE ROLE OF FACIAL APPEARANCE IN GENDER CATEGORIZATION

by

JOSHUA J. SIMPKINS

A thesis submitted in partial fulfillment of the requirements
for the Honors in the Major Program in Sociology
in the College of Sciences
and in The Burnett Honors College
at the University of Central Florida
Orlando, Florida

Spring Term 2011

Thesis Chair: Dr. Shannon Carter

ABSTRACT

Of the many systems of social organization which members of US society use to categorize other members, gender is one of the most important. The gender system operates to place members of US society into categories, and then allocate labor and resources to those members on the basis of their category membership. In order to better understand the gender system, this study examines the methods by which members of US society use the gender system to place other members into a gender category. First, full facial photographs were taken of a group of participants of varying gender, race, ethnicity, and age. Then, parts of each participant's face were isolated digitally and shown to a second group of participants. This second group was asked to identify the sex and/or gender of the individual in the image, indicate how confident they were in this identification, and then write a brief explanation for why they identified the individual in the image as they did.

The analysis conducted by this study supports three findings. First, though the gender categories "male" and "female" are still widely predominant, other categories such as "genderqueer" are seeing use as well. Second, the mouth and lips tend to be seen as more important indicators of gender than other facial features. Finally, while the race and gender category membership of the member doing the categorizing has little or no interaction with the gender categorization process, the race and gender category membership of the member being categorized does have a significant interaction.

DEDICATION

For everyone who isn't easily categorized on the basis of facial appearance;

And for the members, friends, and family of Q House, for supporting me and believing in me even when I didn't.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Shannon Carter, for her patience, her support, and her guidance throughout this project. I would also like to thank the other members of my committee, Dr. Amy Donley, Dr. Elizabeth Grauerholz, and Dr. Maria-Cristina Santana, for their help in editing and developing this paper. I would also like to thank Dr. Jennifer Vest, for her wisdom and her support. Finally, I would like to thank all those who volunteered their faces to create the stimulus images for this project.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1

CHAPTER 2: LITERATURE REVIEW 5

2.1) The Sex Role Behavior Orientation 5

2.2) The Social Psychological Orientation 12

2.3) The Ethnomethodological Orientation 18

CHAPTER 3: THE CURRENT STUDY 23

CHAPTER 4: RESEARCH QUESTIONS AND HYPOTHESES 26

4.1) Question 1: Which Categories? 26

Hypothesis 1 26

Hypothesis 2 26

Hypothesis 3 27

Hypothesis 4 27

4.2) Question 2: Which Parts of the Face? 27

Hypothesis 5 27

4.3) Question 3: Which Social Categories? 28

Hypothesis 6 - 28

CHAPTER 5: METHOD 29

5.1) Phase 1 – Stimulus Image Creation 29

5.2) Phase 2 – Stimulus Image Gender Categorization 32

5.3) Data Analysis 33

CHAPTER 6: SAMPLE 34

6.1) Phase 1 34

6.2) Phase 2 34

CHAPTER 7: DATA ANALYSIS 35

7.1) Gender Categories 35

7.2) Facial Features 39

7.3) Interaction of Other Social Categories 42

CHAPTER 8: DISCUSSION 50

CHAPTER 9: CONCLUSION 55

APPENDIX A: DEMOGRAPHIC SURVEY QUESTIONS 57

APPENDIX A: DEMOGRAPHIC SURVEY QUESTIONS 58

APPENDIX B: STIMULI SURVEY QUESTIONS 59

APPENDIX B: STIMULI SURVEY QUESTIONS 60

APPENDIX C: MASTER STIMULI IMAGES AND STIMULUS IMAGE EXAMPLE 61

APPENDIX C: MASTER STIMULI IMAGES AND STIMULUS IMAGE EXAMPLE 62

REFERENCES 65

LIST OF TABLES

Table 1: Participants' Category Usage in Self-Categorization 35

Table 2: Participants' Category Usage in Target Categorization 36

Table 3: Stimulus Image Confidence Levels and Match Percentages 40

Table 4: Master Stimuli Confidence Levels and Match Percentages 43

Table 5: Gender Group Confidence Levels and Match Percentages 45

Table 6: Age Group Confidence Levels and Match Percentages 46

Table 7: Race Group Confidence Levels and Match Percentages 48

CHAPTER 1: INTRODUCTION

When examining how members of US society interact with one another, researchers have observed a tendency for members to associate other members together into groups or categories (Crawley, Foley and Shehan 2008; McKone, Aitkin, and Edwards 2005; Quinn and Macrae 2005; Ridgeway and Correll 2004). On the basis of their membership in a certain group, members of society can be organized, producing a way to consistently divide labor, responsibility, and finite resources (Lorber 1991). The methods by which these members are placed into groups and then allocated labor and resources on the basis of group membership form intersecting social systems which create and recreate inequality (Crawley et al. 2008; Frye 1983; Ridgeway and Smith-Lovin 1999; Risman 2004; Stokoe 2006).

Of the many systems of inequality which operate within US society, the gender system is one of the most important and ubiquitous (Hawkesworth 1997; Ridgeway 2009; Ridgeway and Correll 2004; Risman 2004; West and Zimmerman 1987; West and Fenstermaker 1995). Three characteristics of the gender system create this importance: the use of gender in initial automatic social perception; the perceived biological basis of gender categories; and the use of dichotomous gender categories.

First, research conducted by cognitive psychologists has indicated that gender category membership is one of the social characteristics automatically and unconsciously attended to by members of US society when encountering another member of society (Irmen 2006; Quinn and Macrae 2005; Stangor et al. 1992). When socially perceiving another individual, members of US

society attend to three basic category memberships - race, age, and gender - as these characteristics are believed to be important indicators of an individual's attributes, attitudes, and behaviors (Irmen 2006; Quinn and Macrae 2005; Ridgeway and Correll 2004). Although other category memberships such as occupation, class, or familial relation may be treated as more salient to the current interaction, gender acts as a kind of "background category" which informs the perceived individual's placement in other categories (Ridgeway and Correll 2004). For example, although a patient in an emergency room may attend to another individual's occupational category membership (i.e., as a "doctor", "nurse", or "receptionist") as more salient than gender due to the context of the interaction, the perceived individual's gender category membership still informs how the occupational membership is interpreted (i.e., as a "female doctor" or a "male nurse").

Second, the gender system in the US links gender category membership to the biological characteristics of an individual (Fausto-Sterling 2000; Kessler and McKenna 1987). It is a social and legal imperative that newborn members are categorized as either male or female shortly after birth (Mealey 2000). Most commonly, this is accomplished by the attending physician's decision to classify an infant's external genitalia as a penis or a vagina (Fausto-Sterling 2000; Kessler 1990). In the case of intersex newborns, individuals with a medical condition that can result in ambiguous genitalia at birth, other biological criteria such as chromosomal sex or fetal gonad sex are looked to as holding the "truth" of an individual's gender category (Fausto-Sterling 2000; Mealey 2000). Since an individual's biology is seen as an innate, natural, essential, and unchanging characteristic of that individual, gender category membership is

likewise socially believed to be innate, natural, essential, and unchanging (Garfinkel 1967; Fausto-Sterling 2000; Kessler 1990; Kessler and McKenna 1987).

Finally, the gender system in US society tends to allow for membership in only two categories, "male" and "female." Category membership is perceived as dichotomous; membership in the category "male" disqualifies an individual for membership in the category "female" and vice versa. Dichotomous membership results in a systemic emphasis on the differences between categories while minimizing recognition of similarities between categories and variation within categories (Birdwhistell 1970; Frye 1983; Garfinkel 1967; Unger 1979).

The gender system interacts with several other systems of social organization and stratification in US society to create an intersectional system of social hierarchies. Systems of gender, race, ethnicity, class, sexuality, and many other social categories all interact with each other to inform an individual's membership across all social systems (Crawley et al. 2008; Frye 1983; Hawkesworth 1997; Kirk and Okazawa-Rey 2010; Ridgeway and Smith-Lovin 1999; Risman 2004; Stokoe 2006). It is impossible to gain an in-depth understanding of how one of these social systems works to create and recreate social inequality without considering how it interacts with other social systems (Cotera 1997; Kirk and Okazawa-Rey 2010; Lorde 1980; West and Fenstermaker 1995). Therefore, although some have argued that the gender system is the most important system of stratification within US society (Hawkesworth 1997; Ridgeway and Smith-Lovin 1999), examining the gender system in isolation from other systems such as race or class limits understanding of social inequality and perpetuates the marginalization of members

of the oppressed categories of the ignored systems (Cotera 1997; Crawley et al. 2008; Kirk and Okazawa-Rey 2010; Lorde 1980; Oyewumi 1998).

Due to the importance and near universal relevance of the gender system within US society, this study argues that an examination of the methods used within the gender system is vital to an understanding of the nature of inequality within US society. These methods can be divided into two analytic groups: methods which function to place members of society into gender categories and methods which function to allocate labor, resources and responsibility to members on the basis of their category membership (Ridgeway and Smith-Lovin 1999; Risman 2004; West and Zimmerman 1987).

This study examines the first analytic group, those methods which function to place members of society into gender categories. It interrogates the methods used by individual members during their initial visual exposure to another member rather than the methods used by institutions or the methods used to maintain incumbency in a category. Finally, this study limits the resources available to members for the act of categorization, allowing them to use only facial characteristics and cues.

CHAPTER 2: LITERATURE REVIEW

Many studies have explicitly attempted to examine the ways in which members of society classify other members into gender categories. These studies tend to be based on one or more of three theoretical orientations: a Sex Role Behavior Orientation, a Social Psychological Orientation, and an Ethnomethodological Orientation. An examination of these studies, and of their strengths and weaknesses, will provide insight into how this study is informed by previous work on the topics of facial appearance and gender attribution and explain what gaps in the current understanding of these topics this study seeks to address.

2.1) The Sex Role Behavior Orientation

The studies grouped here as having a theoretical basis in "sex role behavior" share a common initial hypothesis on which they base their research. This hypothesis is that there are two "sex roles" (or "gender roles", as "sex" and "gender" are often used interchangeably in these studies) that individual members of society enact on a day-to-day basis: a male role and a female role (Barlow et al. 1979; Birdwhistell 1970; Martin 1998). On the whole, they argue, "men and women in the United States hold and move their bodies differently" (Martin 1998:494). Differences have been widely observed in the way in which men and women, on average, sit, stand, walk, gesture, use facial expressions, talk, and do many other activities (Birdwhistell 1970; Martin 1998; Crawley, Foley, and Shehan 2008). These studies, then,

examine which behaviors are assigned to the male or female role, and how the sex role behaviors enacted by individuals in interaction result in their placement in one or another role.

Ray Birdwhistell's (1970) essay on masculinity and femininity as displays serves as the foundation of many of the studies on sex role behaviors that would follow. Birdwhistell differentiated between three types of different, yet interconnected, sexual characteristics: primary, secondary, and tertiary. Birdwhistell writes:

> It seems methodologically useful to me to distinguish between *primary* sexual characteristics which relate to the physiology of the production of fertile ova or spermatozoa, the *secondary* sexual characteristics which are anatomical in nature, and the *tertiary* sexual characteristics which are patterned social-behavioral in form. These latter are learned and are situationally produced. (P. 42)

Birdwhistell assumes that among members of the species *Homo sapiens*, primary, secondary, and tertiary sexual characteristics are all dimorphic into two sex categories, male and female. However, Birdwhistell points to evidence showing that primary sexual characteristics are often unavailable for the determination of sex in day-to-day interaction, and that secondary sexual characteristics, while more available in day-to-day interaction, are only very weakly dimorphic and therefore not a good indicator of sex. Tertiary sexual characteristics, then, must be the most influential resource for the identification of gender (Birswhistell 1970).

Birdwhistell relays reports from informants from seven different societies that indicate that, although which behaviors are typed as "male" or "female" differ from society to society,

each native informant could distinguish male movement from female movement. As well, each informant reported being able to differentiate “feminine” males and “masculine” females. As an illustration of what certain societies might consider differential male or female behavior, Birdwhistell describes three different patterns of behaviors from American society: differences in positioning of the limbs in relation to the torso, differences in positioning of the pelvis and spine, and differences in eye and eye lid behavior. These behaviors, as well as many others still to be discovered, make up what Birdwhistell terms a society’s “gender display and recognition system” (Birdwhistell 1970:45). However, Birdwhistell also urges readers to remember that “*no position, expression, or movement ever carries meaning in and of itself*”, that the context of the gender display will have an effect on how the display is recognized (Birdwhistel 1970:45, emphasis in original).

Birdwhistell’s essay urges a research program that is fairly similar to that urged by later feminist writers (Kessler and McKenna 1978; West and Fenstermaker 1995). In his concluding paragraphs, Birdwhistell exhorts other researchers to remember that gender behavior is not always related to sexual behavior and that gender display and recognition systems are a product of social construction. Birdwhistell argues several times that the ability to display and recognize gender is not an innate characteristic of *Homo sapiens*, but is rather learned in childhood. As well, Birdwhistell acknowledges that gender display and recognition are modified by individuals' other identities and the context in which display and recognition take place (Birdwhistell 1970:45).

Birdwhistell's distinction between primary, secondary, and tertiary sexual characteristics is a very useful methodological tool for talking about and conceptualizing different levels of analysis within a system of gender display and recognition. As well, the position that "gender displays" are only understandable in relation to other identification displays and the context within which they occur has received wide support (Crawley et al. 2008; Fausto-Sterling 2000; Garfinkel 1967; Goffman 1987; West and Fenstermaker 1995). However, there are two primary points on which the basis of this study diverges from Birdwhistell's arguments.

First, Birdwhistell accepts that there exists a "clear demarcation between the production of ova and spermatozoa in organisms of a bisexual species" (Birdwhistel 1970:40). From this basis arises Birdwhistell's tacit assumption that although the behaviors associated with male gender display and female gender display differ from society to society, the "male/female" dichotomy is natural and universal. Both at the time of Birdwhistell's writing and in more recent years, this "clear demarcation" between ova producers and spermatozoa producers in humans had been critiqued as a social construction within the medical and scientific communities rather than an observation of actual cases (Fausto-Sterling 1992, 2000; Garfinkel 1967; Kessler 1990, 1998; Kessler and McKenna 1978; Nicholson 1998). As well, Birdwhistell fails to examine societies wherein other genders than 'male' or 'female' exist, requiring a more complex system of display and recognition (Kessler and McKenna 1978; Lang 1999; Nanda 2000).

Second, Birdwhistell reports the phenomenon of members of a society being recognizable as feminine males and masculine females and then continues on without

comment. This phenomenon would seem to throw doubt on Birdwhistell's main argument in this essay - that tertiary sexual characteristics are the most important in displaying and identifying a member of society's gender. If a member's behaviors were the determining factor in the identification of their gender, then someone displaying 'feminine' behaviors should be identified as a 'female' and an ova producer; if not, then this discrepancy in the meaning of gender messages should result in, as Birdwhistell fears, the destruction of "the necessary conditions for appropriate mating" (Birdwhistell 1970:46). As well, a member can be identified as a certain gender whether or not behavioral cues are an available resource. Members of US society and others can identify another member as one gender or another through a variety of mediums wherein behavioral cues are either absent or extremely limited, such as through text, photographs, film, and telephone conversations (Bruce et al. 1993; Kessler and McKenna 1987; Quinn and Macrae 2005; Williams and Mendelsohn 2008; Yamaguchi, Hirukawa, and Kanazawa 1995).

David A. Barlow and colleagues' (1979) study on sex role motor behavior provides an example of how Birdwhistell's model is applied in practice. The study sets out with the goal of measuring sex role behavior not through self-reported inventories but through the measurement of behaviors by males and females. The study proposed to test a checklist of "masculine" and "feminine" sex role behavior, developed from the work of Birdwhistell and others and from Barlow's own clinical work with adult transsexuals. Barlow and colleagues decided to collect video of subjects standing, sitting, and walking to examine the frequency with which the different behaviors identified on the checklist occurred.

In order to collect subjects for the study, two college age male raters who knew nothing about the study walked through the campus of the University of North Carolina at Greensboro with an experimenter. Raters were told to select passers-by at random and rate them on a 7-point scale from -3 (extremely masculine) to +3 (extremely feminine). In this way, subjects were selected to fill four groups: a) masculine males; b) feminine males; c) masculine females; and d) feminine females. Twelve white subjects were obtained for each group except feminine males, where only ten subjects were obtained. Interesting to note here is that "slightly more liberal criteria in selecting subjects were used for cells in which sex role behavior was opposite to the actual gender" due to difficulty finding subjects in those groups (Barlow et al. 1979). Once a subject was identified by the raters, the experimenter would approach the subject and ask them to cooperate in a study. Subjects who agreed were taken to a nearby room for videotaping.

In the room, subjects were asked to stand next to a chair in front of a video camera while an experimenter pretended to fill out paperwork. Subjects were told that the video camera was turned off and that they would need to calibrate it by walking to a certain spot on the floor. After walking to the spot and back to the chair once, subjects were asked to repeat the procedure, and then asked to take a seat. Subjects were then told about the purpose of the study and informed that they would be videotaped sitting, walking, and standing in front of the video camera. Subjects repeated the same series of behaviors (i.e., walking to the spot on the floor twice, sitting and standing) but this time aware of the purpose of the study and of being videotaped.

The video tapes produced by this process were then shown to two naïve raters who were asked to identify certain behaviors on the checklist developed previously by Barlow. Barlow's checklist identified seventeen feminine behaviors and sixteen masculine behaviors. In this way, a scale from zero to seventeen was developed for feminine behaviors ("0" being no feminine behaviors and "17" being sixteen feminine behaviors), and a scale from zero to sixteen for masculine behaviors ("0" being no masculine behaviors and "16" being sixteen masculine behaviors).

The study concluded that the checklist developed by Barlow would accurately differentiate between males and females, and between masculine females and feminine females, but not between masculine males and feminine males. Barlow and colleagues postulated that the failure of the checklist to differentiate between masculine males and feminine males was likely due to deficiencies in the ability of the first two raters.

This study provides a good example of the methodological weaknesses of the sex role behaviors model of gender promoted by Birdwhistell. First, the study only examined subjects who were identified as demonstrating highly masculine or highly feminine sex roles. If behavioral cues are the most important display of gender, then the proper recognition of ambiguous individuals would be when those cues are most relevant (Kessler and McKenna 1978). The choice to only examine the extreme ends of the spectrum can be seen as the logical result of working from a conceptual framework that utilizes two naturally dichotomous sex roles, “masculine” and “feminine.”

Second, the study assumed that the raters in both instances used only behavioral cues to identify the sex role behavior and biological sex of the study subjects. Although raters were naïve as to the purpose of the study, the raters themselves would categorize the subjects according to gender as they perceived them and then expect certain kinds of behavior of them as a result of that categorization (Huart, Corneille, and Becuart 2005; Irmen 2006; Stangor et al. 1992). While the effect that gendered behavioral expectations would have had on the raters' ratings unknown, the study's failure to control for spoken, visual and contextual indicators of sex role which might have primed gender stereotypes and biased perceptions of behavior in an unknown way.

2.2) The Social Psychological Orientation

Instead of focusing on behavioral cues, these studies oriented within social psychology argue that physical appearance, especially the appearance of the face, is the most important criteria by which individuals are categorized as men and women. Common to these studies is the concept of "categorical thinking", a cognitive strategy whereby a social perceiver places a perceived individual into a certain category - in this case, a gender category - on the basis of various physical and social characteristics (Brown and Perrett 1993; Bruce et al. 1993; Burton, Bruce, and Dench 1993; Chronicle et al. 1994; McKone, Aitkin, and Edwards 2005; O'Toole et al. 1998; Quinn and Macrae 2005; Roberts and Bruce 1988; Yamaguchi et al. 1995).

Three categories have been identified as the dominant categories which members of Western society use in social perception: sex, race, and age (Irmen 2006; Quinn and Macrae

2005; Stangor et al. 1992). However, the contents of these categories tend to be fairly dynamic, and placement in one category may affect an individual's position in another (Irmen 2006). For example, the category "woman" can have several subcategories within it, such as "career woman" or "stay-at-home mother", that result from the category of "gender" being modified by the category of "occupation" (Irmen 2006). As well, certain categories may be more important in different cultures and in different social contexts. For example, although sex categorization is considered to be omni-relevant and vitally important to interaction in Western society (Hawkesworth 1997), it is considered to be less important in other, non-Western cultures (Oyewumi 1998).

The point of inquiry for these social psychological studies is how members of society categorize one another into one of two gender categories - male or female. Instead of arguing that this categorization occurs on the basis of dimorphic behavioral roles (or tertiary sexual characteristics), these studies base their inquiry on the assumption that gender categorization occurs on the basis of dimorphic physical appearance (related to secondary sexual characteristics), especially facial appearance (Brown and Perrett 1993; Quinn and Macrae 2005).

The face is chosen as the principal site for investigation for several reasons. First, in Western societies, the face is "usually the first source of information available about a person" (Jackson 1992:3). Second, individuals tend to attend to the appearance of the face because it is a rich resource for categorizing individuals on the basis of race, age and sex, and for determining the familiarity, emotional status, and gaze direction of an interactional partner

(Jackson 1992; Quinn and Macrae 2005; Stangor et al. 1992). Finally, while other features such as body appearance, hair, grooming, clothing, and behavior may change from interaction to interaction, facial appearance tends to stay fairly stable over multiple interactions (Jackson 1992).

Two types of images have been employed by social psychological studies to examine how members of society utilize facial cues to categorize each other on the basis of gender: prototypical face images, and "natural" face images. In both types of images, "superficial" cues such as hair style, hair length, facial hair, facial jewelry, and facial cosmetics are removed either before the photograph is taken or afterwards using image editing techniques (Bruce et al. 1993). These cues are removed because they are seen as having a social-behavioral origin, while these studies explicitly seek to examine only biological phenomena.

To create prototypical face images, researchers first take photographs of many different individuals and then average various measures from each image (Brown and Perret 1993; Bruce et al. 1993; Burton et al. 1993; Yamaguchi et al. 1995). In this way, researchers can create several different prototypical images: an androgynous image (the average of all images), a masculine image (the average of all images of males), and a feminine image (the average of all images of females). For example, Yamaguchi and colleagues (1995) took full color frontal photographs of twenty-six male and twenty-six female undergraduate students, placed the photographs over a wireframe model using a computer, and then averaged the form of each wireframe and the brightness of each of the three primary colors of pixels in each 2-D image.

By comparison, studies that utilize “natural” face images present participants with images of actual individuals instead of images of abstractions (Bruce et al. 1993; Chronicle et al. 1994; Roberts and Bruce 1988). To create natural face images, researchers simply photograph an individual. Then, the researchers manipulate the photographs or occasionally the participants themselves (Chronicle et al. 1994) to impede sex categorization in some way. For example, in the study conducted by Bruce and colleagues (1993), photographs were taken of individuals, and then the experimental feature was covered with a black rectangle.

Both types of images have been used by studies examining the relative importance of different facial cues in sex categorization, and both types of images have strengths and weaknesses. Prototypical images allow for the importance of a particular facial feature to be examined in the context of a whole face by allowing for the easy swapping of specific facial features between prototypes. By comparison, to examine the importance of particular facial feature with natural face images, the feature must be covered up, usually by a black rectangle, creating images that look quite unlike anything a participant would see in the everyday process of sex categorization. However, a similar critique can be made of prototypical images - prototypical images are completely manufactured. While they may appear closer to what a participant would encounter in everyday life, they are abstractions, not concrete images.

However, although the methodology utilized by social psychological studies serves as a useful model, examining how that methodology is put into practice reveals several blind spots in the previous literature. First, studies tended to either explicitly only use Caucasian faces as stimuli (Chronicle et al. 1994), only used faces from one nationality (Yamaguchi et al. 1995), or,

more commonly, did not report the race, ethnicity or nationality of the individuals used to create the images in their study (Brown and Perret 1993; Bruce et al. 1993; Burton et al. 1993; Roberts and Bruce 1988). As well, studies tended to either only used photographs from individuals of one age group (Chronicle et al. 1994; Bruce et al. 1993; Burton et al. 1993; Yamaguchi et al. 1995) or did not report the age group of photographed individuals at all (Brown and Perret 1993; Roberts and Bruce 1988). Using images of individuals with the same age or race eliminates the possibility to examine how other categories interact with sex.

Second, these social psychological studies tend to use small sample sizes, both in terms of the number of individuals used to create images and the number of participants who evaluated the images (Brown and Perret 1993; Bruce et al. 1993; Chronicle et al. 1994). This is particularly damaging in studies which utilized prototypical facial images in their research design, as a small number of faces averaged together could create an inaccurate prototypical face.

Third, by eliminating so-called "superficial" cues, these studies remove what could be a significant feature of sex categorization in everyday life and create images with no real correlates in everyday life. As well, by declaring these cues social and therefore "not for study", researchers fail to recognize how social behaviors and individual biology interact with each other to create physical appearance (Crawley et al. 2008; Fausto-Sterling 2000). Ironically, several studies identified "eyebrows" as a feature to be experimentally examined (Brown and Perret 1993; Burton et al. 1993; Yamaguchi et al. 1995) and therefore not superficial and social

in nature despite gendered expectations surrounding the maintenance and proper shape of eyebrows (Kessler and McKenna 1978).

Finally, when asking participants to categorize an image into a sex category, most studies only gave participants the option to categorize individuals as either "male" or "female" (Bruce et al. 1993; Roberts and Bruce 1988; Yamaguchi et al. 1995). Forcing participants to make an either-or choice eliminated the possibility for participants to utilize a sex categorization process different from that of the researcher (Crawley et al. 2008).

In concluding this section, it is important to note that several of the studies examined here reported either conflicting or mixed results. Brown and Perret (1993) reported that "all features except the nose carried information about gender when seen in isolation" (p.839) while Chronicle and colleagues (1994) reported that participants made highly accurate sex categorizations using *only* the nose. Bruce and colleagues (1993) reported that information about the sex of the face is conveyed in part by "configural relationships between features" (p. 150), conflicting with the results of Chronicle and colleagues. Yamaguchi and colleagues (1995) reported that eye brows and jaw line were important features, but also reported that this might be because the protypical male face used in their study was "more masculine" (p. 574) than the Caucasian faces used in previous studies due to its darker eyebrows. These ambiguous results indicate that however the gender of a face is determined, it likely involves more than one particular feature or judgment process.

2.3) The Ethnomethodological Orientation

Unlike the studies discussed so far, studies with a theoretical orientation within ethnomethodology explicitly examine how members of society come to a "correct" categorization instead of measuring whether or not that categorization is, in fact, correct (Kessler and McKenna 1978). By changing the focus of research from the "correctness" of a categorization to how that sense of "correctness" is achieved by members of society, ethnomethodological studies focus on how gender categorization is "done" (1978). The overlay study reported by Kessler and McKenna (1978) in their book *Gender: An Ethnomethodological Approach* is an explicitly ethnomethodological study into how members of society place other members into a gender category.

Kessler and McKenna's (1978) *Gender: An Ethnomethodological Approach* builds on the work of Garfinkel's (1967) famous case study of Agnes. Building on Garfinkel's formulation of the natural attitude towards gender, Kessler and McKenna introduce the concept of "gender attribution" as the "process by which one classifies another as male or female" (1978:ix).

Gender attribution involves two steps. First, there is the initial gender attribution (what this study calls "gender categorization") which involves the initial decision to classify another member of society as either a male or a female in the initial stages of interaction. Two areas of self-presentation contribute to the initial gender attribution: 1) "general talk", or what is said and how it is said; and 2) "public physical appearance", or the way in which an individual uses physical accessories and behavior to signal a certain attribution (1978:127). Second, there is the

maintenance of gender attribution over time which is the processes by which members of society maintain their claim to membership as either a male or a female. Similarly, two areas of self-presentation contribute to the maintenance of a gender attribution over time: 1) “the private body”, or the body features hidden from view in most interaction, such as the genitalia or gonads; and 2) “talk about the personal past”, or the ability to provide a self-history that is expected for a member of a particular gender category (1978:127).

On the basis of this model of gender attribution, Kessler and McKenna design a study they called “the overlay study” to examine which body features are necessary for members of society to determine whether another member is male or female. First, they took eleven plastic overlays and drew one physical characteristic or piece of clothing on each. When overlays were placed on top of one another, the result was a human figure with various combinations of stereotypically male or female physical gender characteristics. Of the characteristics, five were stereotypically male (short hair, narrow hips, flat chest, penis, body hair), four were stereotypically female (long hair, wide hips, breasts, vagina), and two were considered “unisex” (“unisex” shirt, “unisex” pants). One final overlay was created as a base, with lines for a torso, arms, and legs, as well as a “non-gender-specific face” (1978:146).

Using these overlays, Kessler and McKenna created 96 different figures with different combinations of features. Each of the 96 figures was then shown to 10 different adults, five males and five females, for a total of 960 study participants. Each participant was asked three questions:

> (1) Is this a picture of a female or a male? (2) Using a scale of 1 to 7, where 1 means not at all confident and 7 means very confident, how confident are you of your answer? (This was, in part, to give us [Kessler and McKenna] information about whether the forced choice in Question 1 was a clear gender attribution or merely a guess.) (3) How would you change the figure to make it into the *other* gender? (1978:146)

Although Kessler and McKenna predicted from previous studies that genitalia would be the defining characteristic for gender attribution when it was visible to participants, they were surprised to discover that the penis appeared to be a much more important cue for a male attribution than the vagina was for a female attribution. In their findings, Kessler and McKenna reported that figures with a penis were attributed as male 96 percent of the time as expected. However, 28 percent of male participants and 43 percent of female participants made an attribution of male to figures with a vagina. They also reported that when asked how they would change a male-attributed figure to make it into the other gender, 38 percent of participants mentioned removing the penis but only 1 percent mentioned adding a vagina; when changing a female to a male, 32 percent of participants mentioned adding a penis, but only 1 percent mentioned removing the vagina.

Along with this apparent importance of the penis, Kessler and McKenna reported participants had a tendency to make a generic male attribution. Figures with mixed gender characteristics that also had their genitalia covered were attributed male by 69 percent of participants. As well, participants weighted female characteristics as less important than male

characteristics. Although figures with a penis were attributed male by participants 96 percent of the time, figures were only attributed female with a similar frequency (95 percent) if they had a vagina along with two or three other female characteristics.

Kessler and McKenna conclude that "in the social construction of gender 'male' is the primary construction" (1978:159). In order to be attributed as male, a member of society only needs to be seen as having one concrete "male" characteristic. Congruently, in order to be attributed as female, a member of society needs to be seen as not having any concrete "male" characteristics. From these conclusions, they propose a cognitive schema for the attribution of gender within Western society: "*See someone as female only when you cannot see them as male*" (1978:158, emphasis in original).

Kessler and McKenna's overlay study produces an amazingly detailed description of how gender categorization occurs between individuals in Western society. As well, their larger work in *Gender: An Ethnomethodological Approach* that both informed and was informed by the overlay study has been hailed as a radical approach to gender decades ahead of its time (Crawford 2000; Hawkesworth 1997). Kessler and McKenna's (1978) theorization of the gender attribution process as a whole along with the overlay study itself informs the theoretical orientation and methodology of the current study. However, there are three relatively minor weaknesses in the overlay study which this study will seek to address.

First, by asking participants "is this a picture of a female or a male," (1978:146) Kessler and McKenna limit the participants' responses to those they already expect. There is no room for the participants to respond outside of the male/female binary. Second, by using a sketch of

a hypothetical person instead of photographs of actual individuals or what appear to be individuals as in the social psychological studies which used prototypical faces, participants are placed one more step away from the actual practice of gender attribution as it occurs in commonplace, daily settings.

Finally, the largest weakness in the overlay study is Kessler and McKenna's failure to take into account the influence of the "non-gender-specific face" and the lines of the arms, legs, and torso common to all of the figures (Kessler and McKenna 1978:146). As the researchers in the social psychological studies above pointed out, the face is a very rich resource used in social interaction (Jackson 1992; Quinn and Macrae 2005; Stangor et al. 1992). As such, participants in the study likely looked to the face of the figures as an important signifier of each figure's gender. If the face and body lines were interpreted by the majority of participants as male characteristics, then that could explain the tendency of participants to make male attributions more commonly than female attributions to mixed characteristic figures.

CHAPTER 3: THE CURRENT STUDY

The purpose of this study is to examine the methods by which members of US society place other members into gender categories at the level of initial interaction. As an improvement upon previous studies on the topic, this study focuses on four main points.

First, this study seeks to examine how members of society make a gender categorization. Second, this study focuses on the use of the face in gender categorization. Third, this study will examine which parts of the face produce a more confident gender categorization, and how the methods for gender categorization change depending on which resources are available. Fourth, this study focuses on how gender categorization is made relative to the gender identities, racial or ethnic identities, age and education level of the member making a categorization and the member being categorized

In order to examine how members of society make a gender categorization, participants were asked to both categorize a stimulus by gender and explain why they made that categorization. According to ethnomethodological theories of gender, members of US society believe that normal, adult members should be able to make a gender categorization easily, correctly and on the basis of unspoken, taken-as-granted "good reasons" for doing so (Garfinkel 1967; Heritage 1984; Kessler and McKenna 1978; Speer 2005). By making gender categorization difficult, allowing for the possibility of an "incorrect" categorization, and calling into question the participant's ability to produce those unspoken, taken-as-granted "good reasons", participants' identities as normal, adult members of society are called into question (Speer

2005). In order to maintain their identity as normal, adult members of society, participants must place stimuli into the "correct" gender category with a high level of confidence and provide acceptable "good reasons" for doing so. Which categories are used and which explanations are considered "good reasons" can then be analyzed for insight into how gender categorization works (Heritage 1984; Speer 2005; Stokoe 2006).

In order to focus on how the face is utilized in gender categorization, facial images were used as stimuli presented to participants for categorization. The face was chosen for three reasons. First, in US society the face is usually the first available source for information about an individual, especially in face-to-face interaction (Jackson 1992). Second, individuals tend to attend to the appearance of the face as a rich resource for categorizing individuals on the basis of gender, race and age, and for determining the familiarity, emotional status and gaze direction of an individual (Jackson 1992; Quinn and Macrae 2005; Stangor et al. 1992). Third, while other features such as body appearance, hair, grooming, clothing, and behavior may change from interaction to interaction, facial appearance tends to stay fairly static over multiple interactions (Jackson 1992). Unlike in the social psychological studies discussed above, the facial images in this study were created from individuals in what they considered their natural, day-to-day appearance.

In order to examine which parts of the face allow members of society to make a more confident gender categorization, stimuli were created to present to participants that only show a portion of a whole face. As well, after being asked to place an individual into a gender

category on only the basis of one part of the face, participants were asked to indicate how confident they were in their categorization.

Finally, in order to examine how gender categorization occurs in relation to a member's incumbency in other social categories commonly used in US society, two practices were used. First, the sample utilized to create the stimuli presented to participants was more diverse with regards to racial identity, age and gender identity than stimuli used in previous studies. Second, data were collected about the age, gender identity, race or ethnic identity and education level for members of both the stimulus sample and the participant sample.

CHAPTER 4: RESEARCH QUESTIONS AND HYPOTHESES

This study sought to answer three research questions about gender categorization. In order to better organize the findings of the study, seven hypotheses were developed as possible answers to these questions.

4.1) Question 1: Which Categories?

What gender categories are used by members of US society in the accomplishment of gender categorization?

Hypothesis 1 - The categories "male" and "female" will be the primary categories. Many writers have argued that in modern US society "male" and "female" are the only two categories which the gender system allows (Fausto-Sterling 2000; Kessler and McKenna 1978; Lorber 1991; Martin 1998; West and Zimmerman 1987).

Hypothesis 2 - Other gender categories - such as "transgender," "genderqueer," or "mixed gender" - will be used rarely, but not disused entirely due to the increased visibility of members of society who fall into one of these categories (Crawley et al. 2008; Devor 1989; Dozier 2005).

Hypothesis 3 - The category "male" and other categories which correspond to "male" will be used disproportionately more often than other categories as a result of Kessler and McKenna's cognitive schema: "*See someone as female only when you cannot see them as male*" (1978:158, emphasis in original).

Hypothesis 4 - A gender category will be applied to most, if not all, of the stimuli presented. This is to say that members will tend to not respond with "I don't know" or "I can't tell," but will rather make a guess and then perform some kind of identity maintenance in the case that their guess was "wrong" (Garfinkel 1967; Kessler and McKenna 1987; Speer 2005).

4.2) Question 2: Which Parts of the Face?

How are particular parts of the face and the relationship between them used in the accomplishment of gender categorization?

Hypothesis 5 - The "rightness" of a gender categorization, and a member's confidence that they have made the "right" categorization, will differs with which part of the face is shown to the member. Previous studies have produced mixed results on which particular parts of the face are more important to the accomplishment of gender categorization, or whether individual parts of the face are even that important without configural information (Brown and Perrett 1993; Bruce et al. 1993; Burton et al. 1993; Chronicle et al. 1994; McKone et al. 2005; O'Toole et al. 1998; Quinn and Macrae 2005; Roberts and Bruce 1988; Yamaguchi et al. 1995)

4.3) Question 3: Which Social Categories?

How does an individual's membership in other social categories affect their categorizations of others? How does a target's membership in other social categories affect categorizations about them?

Hypothesis 6 - The "rightness" of gender categorizations and the confidence with which they are made will differ depending on the category memberships of the individual being categorized. Different members will demonstrate different gender cues to differing degrees, which will make them easier or harder to "rightly" categorize. Likewise, a member's confidence in their categorizations will be affected by these same category memberships. Gender, age, and race are the three category memberships established during automatic, initial categorization (Irmen 2006; Quinn and Macrae 2005; Stangor et al. 1992). Since members categorize targets on the basis of gender, age, and race almost simultaneously, a target's membership in an age or race category likely has an effect on gender categorization and vice versa. Also, since this initial categorization involves the identification of targets as being "the same as" or "different than" the member doing the categorizing, the member's category memberships likely also has an effect on how they categorize others (Quinn and Macrae 2005; Ridgeway 2009).

CHAPTER 5: METHOD

Data were collected using a two-phase survey method. In Phase 1, facial images were taken of 10 participants, matched with those participants' demographic survey responses, and then selectively cropped to make categorization stimuli. In Phase 2, those stimuli were presented through an internet survey to other participants.

5.1) Phase 1 – Stimulus Image Creation

First, stimulus images to present to participants in Phase 2 were created. Drawing from the investigator's personal contacts, 74 individuals were invited to participate in the study through an event posted on the social networking site Facebook. Participants were asked to arrive at the sampling location, the investigator's personal residence, having dressed and prepared their faces as they would normally do during their day-to-day lives.

Upon arriving at the sampling location, participants was asked to complete a short online survey, created through the SurveyGizmo survey software, which informed them of the purpose of the study and what they would be asked to if they agreed to participate. The survey collected information about each participant's age, gender identity, race/ethnicity, and education level (see Appendix A). After completing the survey, participants were assigned an "Image ID Number" to later associate survey responses with participant photographs. Participants were then photographed sitting in a chair in front of a white, neutral surface using a Samsung S630 digital camera secured on a tripod. Distance from the camera to the

participant, lighting, and camera angle were all kept as similar as possible for all participants. Three facial photographs were taken of each participant.

From the original pool of 74 individuals, 14 were able to participate in the study. A secondary sample of 10 participants was selected from these 14. One photograph of each member of this secondary sample was then chosen to create the images used in Phase 2, resulting in a final sample of 10 master images (See Appendix C).

These 10 master images were then edited, using the Paint.NET graphics editor, into the images used in Phase 2. Master images were manipulated in three ways before being edited further. First, every master image was cropped to center the image on the face of the participants. Second, 4 images (1, 3, 6 and 8) were resized so that each image was the same size, 1050 by 1200 pixels. Third, two images, 7 and 10, were edited to remove red-eye. Except for these three cases, none of the master images were retouched, recolored, edited or manipulated in any way so as to keep them as close as possible to the facial photographs or actual faces a member of society might encounter in their normal, day-to-day interactions.

Using Paint.NET, a set of 15 stimulus images was created for each of the 10 master stimuli. Three of the stimulus images - the “facial oval,” “facial oval-negative,” and “full facial” images - were chosen to provide a broad reading of the gender category of a face, and to recreate images similar to those used in previous studies (Brown and Perrett 1993; Huart et al. 2005; McKone et al. 2005; Yamaguchi et al. 1995). Each of the other 12 stimulus images served one of two purposes. “Limited” images showed participants one facial feature without facial context to examine the role of specific features in gender categorization. “Full” images focused

a participant's attention on a central feature, but also provided contextual information from the area around the feature.

Each set of 15 stimulus images contained:

1) a "facial oval" image, showing the brow, eyes, nose, mouth, and chin without the hair, ears, neck or background;

2) a "facial oval - negative" image, showing the hair, ears, neck and background without the brow, eyes, nose, mouth and chin;

3) a "brow - limited" image, showing the brow;

4) a "brow - full" image, showing the brow in a horizontal stripe limited vertically at the mid-forehead and eyes;

5) an "eyes - limited" image, showing the eyes;

6) an "eyes - full" image, showing the eyes in a horizontal stripe limited vertically at the top and bottom of the eye lashes;

7) a "nose - limited" image, showing the nose;

8) a "nose horizontal - full" image, showing the nose in a horizontal stripe limited vertically at the mid-forehead and upper lip;

9) a "nose vertical - full" image, showing the nose in a vertical stripe limited horizontally at the edge of the nostrils

10) a "mouth - limited" image, showing the mouth;

11) a "mouth - full" image, showing the mouth in a horizontal strip limited vertically at the bottom of the nose and the bottom of the lower lip;

12) a “jaw line - limited” image, showing the chin and cheeks, limited so as to not show the ears, mouth or nose;

13) a “lower face - full” image, showing the lower half of the master image, limited vertically at the middle of the nose so as to show the lower half of the nose, the mouth, and the jaw line;

14) an “upper face - full” image, showing the upper half of the master image not shown in the “lower face - full” image;

15) a “full facial” image.

(For examples of these stimulus images, see Appendix C)

After being created, the 150 stimulus images were uploaded into an image library using the SurveyGizmo survey software for use in Phase 2. Images were sorted into “type groups” according to stimulus type - ie, the 10 “facial oval” images made up type group 1, the 10 “facial oval-negative” images made up type group 2, the 10 “brows limited” images made up type group 3, etc - resulting in 15 type groups.

5.2) Phase 2 – Stimulus Image Gender Categorization

In Phase 2, stimulus images created in Phase 1 were used in a survey instrument created through the SurveyGizmo survey software. The instrument was divided into two sections: a demographic section and a stimulus response section. In the demographics section, participants were asked to respond to the same demographic questions presented to participants in Phase 1 (see Appendix A).

In the stimulus response section, participants were presented with 15 stimulus images, 1 chosen randomly from each Type Group. For each stimulus image, participants were asked three questions. First, participants were asked to categorize the person in the image on the basis of sex or gender. Second, participants were asked to indentify, on a scale from 1 to 7, how confident they were in their answer to the first question. Last, participants were asked to explain why they answered the first two questions as they did (see Appendix B).

In order to protect the anonymity of participants, IP addresses and other identifying information was not collected. If a survey participant elected to withdraw from the survey without completing it, their responses were not used in later analysis.

5.3) Data Analysis

The data from Phase 2 was analyzed using Microsoft Excel 2007 and SPSS Statistics 18. Two types of statistical tests were used to analyze the quantitative data. First, Chi-square tests were performed to examine the proportional usage of the "male" category. Second, one-way analysis of variance (or one-way ANOVA) tests were performed to examine how different parts of the face or different social category memberships affected gender categorization.

CHAPTER 6: SAMPLE

6.1) Phase 1

In Phase 1, participants were recruited through a snowball sample via Facebook (N = 14). From these initial participants, a final group of participants was chosen (N = 10) for the creation of the stimulus images used in Phase 2. In this final group, 7 participants identified as female and 3 as male; 1 identified as Asian or Asian American, 2 as Black or African American, 1 as Hispanic or Latino, 5 as Non-Hispanic White, and 1 did not respond. The age of participants ranged from 20 years old to 77 (M = 33.4, SD = 17.73).

6.2) Phase 2

For Phase 2, participants were recruited though a snowball sample via Facebook (N = 306). Of the participants, 206 (67.3 percent) were coded as "female," 89 (29.0 percent) were coded as "male," and 11 (3.6 percent) were coded as "Other." Non-Hispanic White was the most represented race in the sample at 237 (77.5 percent) participants. Participants ranged in age from 18 to 67 years old (M = 29.05, SD = 10.81).

CHAPTER 7: DATA ANALYSIS

In this data analysis, only the quantitative data collected were analyzed.

7.1) Gender Categories

Participants displayed their perceptions of gender categories in two places: in their self-categorization of gender, and in their categorization of the gender of the stimulus images.

Table 1: Participants' Category Usage in Self-Categorization

Category	Count	Percentage
---------- Male/Female ----------		
Female	199	65.03 %
Male	86	28.10 %
Total	***285***	***93.14 %***
---------- Do Correspond to Male or Female ----------		
Girl	2	.65 %
Femme	1	.33 %
Female/Feminine	1	.33 %
Lady	1	.33 %
Masculine	1	.33 %
Woman	1	.33 %
Woman-Identified	1	.33 %
Male Cross Dresser	1	.33 %
Heterosexual Male Cross Dresser	1	.33 %
Total	***10***	***3.27 %***
---------- Do Not Correspond to Male or Female ----------		
Male to Female Transsexual	1	.33 %
Male to Female Transgender	1	.33 %
Transwoman	1	.33 %
Trans	1	.33 %
MTF Genderqueer	1	.33 %
Gender Fluid	1	.33 %
Genderqueer	1	.33 %
Total	***7***	***2.29 %***

Category	Count	Percentage
---------- *Other* ----------		
Born Male	1	.33 %
Natal Male	1	.33 %
N/A	1	.33 %
Total	***3***	***1.00 %***

In their self-categorization, 285 participants identified their gender as either explicitly "male" or "female." The next most commonly used categories (N = 8) correspond to "male" or "female" at least slightly, such as "Woman-identified," "masculine," "femme," and "girl." As well, two participants used the gender categories "heterosexual male crossdresser" and "Male Cross Dresser" which can be understood as sub-categories within the category "male."

Of the remaining 11 participants, 6 used a category which included the terms "trans," "genderqueer," or a combination of the two. The categories "gender fluid" and "intersexed" were each used by 1 person. Finally, of the 3 last participants, 2 identified the category they were placed in by others at birth ("born male" and "natal male"), and 1 refused to answer at all ("N/A").

Table 2: Participants' Category Usage in Target Categorization

Category Code	Count	Percentage
Female	2524	55.44 %
Male	1907	41.88 %
Other - IDK	67	1.47 %
Other - No Gender	47	1.03 %
Other - Queer	8	.18 %
Total	***4553***	***100.00 %***

Over the course of the study, participants made a total of 4553 total categorizations of stimulus images. Of these, 4347 (95.48%) categorizations explicitly used the categories "male" or "female." The next most commonly used categories (N = 83) were categories which correspond to "male" or "female" at least slightly. Gender categories which did not correspond to "male" or "female" were used by 7 respondents a total of 8 times.

Out of the 306 participants, 16 (5.23 percent) used at least one gender category which did not correspond to the categories "male" or "female." 303 participants (99.35 percent) used the category "female" or a category which corresponded to it, and 302 (99.02 percent) used the category "male" or a category which corresponded to it. Thus, it was concluded that "male" and "female" were the primary categories used by participants; other gender categories were used by a small percentage of respondents.

The remaining 114 categorizations were split into two groups. 67 responses were refusals to make a definite categorization because the respondent felt there was insufficient information available to make a categorization. This includes responses like "ambiguous," "I can't tell," "I don't know," or "?". As well, 47 responses from 4 participants indicated that it was impossible to determine the gender of an individual from physical appearance alone. Interestingly, 2 of these 4 participants argued that it was impossible to categorize gender from physical appearance alone when presented with some stimulus images, but made a gender categorization when presented with other stimulus images.

Out of the 306 participants, 31 (10.13 percent) responded at least once to being asked to provide the "sex/gender" of a target with an answer which allowed them to avoid making a

gender categorization. As well, out of the 4553 total responses collected, 114 (2.50 percent) responses avoided making a gender categorization. Thus, it was concluded that participants in this study applied a gender category to a stimulus image in most cases.

Of the 4439 images that were categorized, 3151 (70.98 percent) were of master stimuli who self-categorized as "female" and 1288 (29.02 percent) were of master stimuli who self-categorized as "male." This ratio of "female" to "male" images does not differ significantly from the expected ratio based off of the ratio of male to female self-categorized master stimuli (χ^2 = 2.049, p = .152). These quantities are used as the expected quantities in the following chi-square tests.

Participants in the study made gender categorizations in 4439 instances. Of these, 2524 (56.86 percent) instances used the category "female" or a category which corresponds to female and 1907 (42.96 percent) instances used the category "male" or a category which corresponds to male. These quantities differ significantly from the expected quantities (χ^2 = 422.689, p < .000). In 3456 instances, the categorization made by the participant matched the self-categorization made by the master stimuli in the image. Of these matching categorizations, 2337 (68.02 percent) were matched as "female" and 1099 (31.98 percent) were matched as "male." These quantities differ significantly from expected quantities (χ^2 = 14.663, p < .000). Finally, in 1003 instances, the categorization made by the participant did not match the self-categorization made by the master stimuli in the image. Of these mismatched categorizations, 187 (18.64 percent) were mismatched as "female" and 808 (80.56 percent) as "male." These quantities also differ significantly from the expected quantities (χ^2 = 1315.522, p < .000). Thus,

it was concluded that the category "male" was used disproportionately more often than other categories.

7.2) Facial Features

In order to examine how different parts of the face affect the "rightness" of a gender categorization, participant's categorizations of stimulus images were coded into 5 codes. Responses which explicitly used the category "male" or a category which corresponds to "male" - such as "masculine" or "man" - were coded as "Male." Similarly, responses which explicitly used the category "female" or a category which corresponds to "female" - such as "feminine" or "woman" - were coded as "Female." Responses which used categories which did not explicitly use or correspond to "male" or "female" - such as "trans" or "genderqueer" - were coded as "Other - Queer." Finally, responses which indicated an inability to make a categorization were coded as "Other - IDK" and responses which argued that it was impossible to categorization from physical appearance were coded as "Other - No Gender."

For several of the following tests, responses coded as "Other - IDK" and "Other - No Gender" were removed from the sample. This was done for two reasons. First, both of these types of response can be read as the participant saying "I am unable to make a categorization on the basis of the information provided." Other research tends to suggest that members of US society in such a situation do one of two things - they either make a way to discover more information about the person in question to learn their "true" gender, or they make a tentative gender categorization and move forward tentatively (Frye 1983; Irmen 2006; Kessler and

McKenna 1978; Stangor et al. 1992). Since the survey instrument was unable to test for these or other actions, it is irresponsible to make assumptions about what participants "really meant" in these cases. Secondly, responses coded as "Other - IDK" lack the ability to be accurately matched for the "rightness" of the gender categorization. As such, these responses will be excluded from tests of "rightness," but included on tests of confidence intervals. When included in these tests, responses coded "Other - IDK" will be assigned a confidence interval of "1."

Table 3: Stimulus Image Confidence Levels and Match Percentages

Stimulus Image ID	Mean Confidence Level[a]	Mean Match Percentage[b]
Mouth - Limited	4.60	.79
Nose - Limited	3.19	.51
Eyes - Limited	3.67	.63
Brows - Limited	3.86	.66
Eyes - Full	4.54	.76
Mouth - Full	5.36	.90
Brows - Full	4.46	.74
Nose - Full - V	5.17	.82
Jaw Line - Limited	4.26	.73
Nose - Full - H	5.16	.80
Facial Oval - N	5.36	.76
Lower Half	5.65	.85
Upper Half	5.26	.82
Facial Oval	5.77	.93
Full Facial	6.04	.90
Total	***4.82***	***.77***

Notes

a: "Mean Confidence Level" is the average confidence level on a scale from 1 ("Not Confident at all") to 7 ("Completely Confident").

b: "Mean Match Percentage" is the percentage of responses where participants' categorizations matched the self-categorization of the individual in the image.

To examine how the rate at which participants' gender categorizations matched the master stimuli's self-categorization varied depending on the part of the face shown to participants, a one-way ANOVA was performed. The data set tested excluded responses coded as "Other - No Gender" and "Other - IDK." From this test it was concluded that the ability to "rightly" gender categorize a target - the match percentage - varies relative to the facial feature available ($F = 21.943$, $p < .000$).

Overall, participants categorized a target "rightly" with good success (M = .77). Categorizations matched the least on stimulus image 7, the "nose - limited," image (M = .51) and the most on stimulus image 1, the "facial oval" image (M = .93). Of the five "limited" images, the mouth had the highest match percentage (M = .79), significantly higher than the other limited images ($p < .05$). Of the "full" images, the mouth had the highest match percentage (M = .90), significantly higher than the eyes and brows images ($p < .05$), but not the nose-horizontal and nose-vertical images. Interestingly, the "mouth - full" image had a significantly higher match percentage than the "jaw line - limited" image ($p < .000$) despite the fact that the images tended to show similar areas on the face. This implies that the lips are seen as a better indicator of gender category than the jaw line or beard hair.

To examine how participants' confidence levels varied depending on the part of the face shown to them, a one-way ANOVA was performed. The data set tested excluded responses coded as "Other - No Gender." From this test it was concluded the confidence a member has in the "rightness" of their categorizations varies relative to the facial feature available ($F = 73.026$, $p < .000$).

Overall, participants tended to have moderate levels of confidence in the "rightness" of their categorizations (M = 4.82). Participant confidence was lowest on stimulus image 7, the "nose - limited" image (M = 3.19) and highest on stimulus image 15, the "full facial" image (M = 6.04). Of the "limited" images, the mouth images had the highest average confidence (M = 4.60), significantly higher than the other limited images ($p < .05$). Of the "full" images, the mouth had the highest average confidence (M = 5.36), significantly higher than the eyes and brows images ($p < .05$), but not the nose-horizontal and nose-vertical images.

A participant's confidence level correlated significantly with whether or not their categorization matched ($r = .311$, $p < .000$). Participant confidence and ability to "rightly" categorize gender tended to increase as the amount of the face available to them for categorization increased, with two notable exceptions: the "mouth - limited" image and the "mouth - full" image.

Confidence level ($r = .354$, $p < .000$) and match percentage ($r = .174$, $p < .000$) also correlated significantly with the percentage of the full facial image available to the participant. However, there were three apparent outliers - the "mouth - limited," "mouth - full," and "jaw line - limited" images -that when removed increased the significance of the correlation for both confidence level ($r = .446$, $p < .000$) and match percentage ($r = .241$, $p < .000$).

7.3) Interaction of Other Social Categories

Before examining how different social category memberships effect gender categorization, the mean confidence levels and match percentages for each of the 10 master

stimuli were calculated and compared using a one-way ANOVA test. From this test it was concluded that both confidence level and match percentage varies between the 10 master stimuli ($F = 57.483$, $p < .000$; $F = 43.289$, $p < .000$).

Table 4: Master Stimuli Confidence Levels and Match Percentages

Master Stimuli ID	Mean Confidence Level[a]	Mean Match Percentage[b]
1	4.27	.68
2	5.59	.90
3	4.08	.60
4	4.63	.71
5	5.07	.91
6	4.76	.70
7	4.78	.85
8	4.90	.83
9	6.11	.96
10	4.20	.64
Total	***4.82***	***.76***

Notes

a: "Mean Confidence Level" is the average confidence level on a scale from 1 ("Not Confident at all") to 7 ("Completely Confident").

b: "Mean Match Percentage" is the percentage of responses where participants' categorizations matched the self-categorization of the individual in the image.

Pairwise comparisons of the 10 master stimuli revealed participants reported the highest mean confidence levels and had the highest match percentages when categorizing master stimuli 2 (M = 5.59; M = .90), master stimuli 5 (M = 5.07; M = .91) and master stimuli 9 (M = 6.11; M = .96). Participants reported the lowest mean confidence levels and lowest match percentages when categorizing master stimuli 1 (M = 4.27; M = .68), master stimuli 3 (M = 4.08; M = .60), and master stimuli 10 (M = 4.20; M = .64).

In order to examine how the membership of participants and master stimuli in various social categories affect the confidence with which a gender categorization is made and the "rightness" of that categorization, participant and master stimuli gender, age, and race/ethnicity were coded.

Self-categorization of gender for both groups was coded using the same method used for coding participants' categorization of stimulus images above. Due to the disproportionate number of participants under the age of 26, age was coded into 2 codes: "18 to 26 years old" (N = 180) and "27 and older" (N = 126). Master stimuli age was coded using the same method. Finally, because more than twice as many participants identified their race/ethnicity as "Non-Hispanic White" than every other group combined, participant race was recoded as well. Participants who identified themselves as "Non-Hispanic White" were coded as "White" (N = 237) and all other participants were coded as "Non-White" (N = 55). Participants who responded "Prefer Not To Respond" were not included in the data set when testing for differences on the basis of race (N = 14). Master stimuli race was coded using the same method.

In each of the test scenarios described below, responses are coded into groups using the nomenclature "group category one x group category two." This indicates that this group is made up of data points where the participant was coded into group category one and the master stimuli was coded into group category two. For example, the group "Female x Male" would contain all data points where the participant was coded as "Female" and the master stimuli was coded as "Male."

Table 5: Gender Group Confidence Levels and Match Percentages

Participants' Category	x	Master Stimuli Category	Confidence Level		Match Percentage	
			Mean	*Std. Dev.*	*Mean*	*Std. Dev.*
---------- All Participants ----------						
All	x	Female	4.67	1.832	.76	.430
All	x	Male	5.50	1.821	.86	.344
---------- All Master Stimuli ----------						
Female	x	All	4.84	1.855	.77	.419
Male	x	All	4.83	1.801	.79	.410
---------- Pairwise ----------						
Female	x	Female	4.57	1.844	.74	.441
Female	x	Male	5.50	1.821	.86	.344
Male	x	Female	4.54	1.766	.76	.426
Male	x	Male	5.52	1.698	.85	.360

In order to examine the interaction of participants' and target master stimuli's gender in gender categorization, data points were coded into four groups: Female x Female, Female x Male, Male x Female, and Male x Male. Respondents whose gender was coded as "Other" were excluded from this data set due to their small number (N = 148) compared to the Female x Female grouping. Two one-way ANOVA tests were then performed on this data set, one for confidence levels and one for match percentages. From these tests it was concluded that both confidence level ($F = 81.356$, $p < .000$) and match percentage ($F = 23.652$, $p < .000$) vary relative to the gender makeup of the categorizing/categorized pair.

Pairwise comparisons of the four groupings revealed that the mean confidence levels and match percentages for the Female x Female and Male x Female groups were not significantly different ($F = .106$, p = .745; $F = 2.549$, $p = .110$), nor were the mean confidence

levels and match percentages for the Female x Male and Male x Male groups (F = .057, p = .811; F = .538, p = .463). However, the mean confidence levels and match percentages of the Female x Female and Male x Female groups were significantly different from the confidence levels and match percentages of the Female x Male and Male x Male (F = 244.009, p < .000; F = 67.706, p < .000). Thus, it was concluded that while confidence level and match percentage does not vary significantly relative to the gender category of the participant, they do vary relative to the gender category of the master stimuli. Participant confidence and match percentage were significantly higher when the master stimuli being categorized was male.

Table 6: Age Group Confidence Levels and Match Percentages

Participants' Category	x	Master Stimuli Category	Confidence Level		Match Percentage	
			Mean	*Std. Dev.*	*Mean*	*Std. Dev.*
---------- All Participants ----------						
All	x	18 to 26	4.71	1.790	.78	.417
All	x	27 and Older	4.93	1.912	.77	.419
---------- All Master Stimuli ----------						
18 to 26	x	All	4.79	1.854	.78	.417
27 and Older	x	All	4.97	1.809	.76	.430
---------- Pairwise ----------						
18 to 26	x	18 to 26	4.70	1.819	.78	.413
18 to 26	x	27 and Older	4.88	1.943	.78	.417
27 and Older	x	18 to 26	4.74	1.746	.77	.423
27 and Older	x	27 and Older	5.00	1.856	.77	.424

In order to examine the interaction of participants' and target master stimuli's age in gender categorization, data points were coded into four groups: 18 to 26 x 18 to 26, 18 to 26 x 27 and older, 27 and older x 18 to 26, and 27 and older x 27 and older. Two one-way ANOVA tests were then performed on this data set, one for confidence levels and one for match

percentages. From these tests it was concluded that while match percentage does not vary relative to age makeup of the categorizing/categorized pair ($F = .359$, $p = .783$), confidence level does vary relative to age makeup ($F = 5.948$, $p < .000$).

Pairwise comparisons of the four groupings revealed two things. First, participants ages 18 to 26 reported confidence levels significantly lower than participants ages 27 and older when categorizing master stimuli in the same age group as the participant (ie, 18 to 26 categorizing 18 to 26 and 27 and older categorizing 27 and older) ($F = 15.040$, $p < .000$). Second, participants ages 27 and older reported significantly higher confidence levels when categorizing master stimuli in the 27 and older age group than when categorizing master stimuli in the 18 to 26 age group ($F = 9.837$, $p = .002$). Participants ages 18 to 26 also reported significantly higher confidence levels when categorizing master stimuli ages 27 and older than when categorizing master stimuli ages 18 to 26 although the significance of this difference was weaker than previous differences ($F = 6.007$, $p = .014$). Thus, it was concluded that while match percentage does not vary significantly relative to the age makeup of a categorizing/categorized pair, confidence level does. Participants in both age groups reported significantly lower confidence levels when categorizing master stimuli in the 18 to 26 age group.

Table 7: Race Group Confidence Levels and Match Percentages

Participants' Category	x	Master Stimuli Category	Confidence Level		Match Percentage	
			Mean	*Std. Dev*	*Mean*	*Std. Dev.*
---------- All Participants ----------						
All	x	White	5.05	1.844	.81	.393
All	x	Non-White	4.51	1.804	.72	.450
---------- All Master Stimuli ----------						
White	x	All	4.79	1.854	.78	.417
Non-White	x	All	4.97	1.809	.76	.430
---------- Pairwise ----------						
White	x	White	5.01	1.859	.82	.388
White	x	Non-White	4.48	1.804	.72	.450
Non-White	x	White	5.22	1.770	.78	.412
Non-White	x	Non-White	4.60	1.804	.72	.450

In order to examine the interaction of participants' and target master stimuli's race in gender categorization, data points were coded into four groups: White x White, White x Non-White, Non-White x White, and Non-White x Non-White. Two one-way ANOVA tests were then performed on this data set, one for confidence levels and one for match percentages. From these tests it was concluded that both confidence level and match percentage vary relative to the racial makeup of the categorizing/categorized pair ($F = 32.529$, $p < .000$; $F = 17.103$, $p < .000$).

Pairwise comparisons of the four groups revealed two things. First, there was not a significant difference between reported confidence levels of White and Non-White participants when categorizing White ($F = 5.462$, $p = .020$) or Non-White ($F = 1.104$, $p = .293$) master stimuli. However, mean reported confidence levels were significantly higher for White master stimuli than for Non-White master stimuli ($F = 90.857$, $p < .000$). Second, there was not a significant

difference between match percentages of White and Non-White participants for White (F = 2.536, p = .111) or Non-White (F = .002, p = .968) master stimuli. However, mean match percentage was higher for White master stimuli than for Non-White master stimuli for both White (F = 45.613, p < .000) and Non-White (F = 4.608, p < .032). Thus, it was concluded that while confidence level and match percentage do not vary significantly relative to the race category of a participant, they do vary relative to the race category of the master stimuli.

CHAPTER 8: DISCUSSION

The results of this study serve as an exploration of what Kessler and McKenna termed the "rules" of gender categorization (1978:158). They also serve to test the arguments put forth by Kessler, McKenna, and others about the natural attitude with regards to gender - the basic, unquestionable axioms which undergird intersubjective social reality (Garfinkel 1967; Kessler and McKenna 1978; West and Zimmerman 1987). By finding answers to the research questions posed here, this study builds understanding of the gender system as it exists today.

In answering the question "What gender categories are used by members of US society in the accomplishment of gender categorization?", this study sought to test four hypotheses.

The first hypothesis was that members of US society will tend to use the categories "male" and "female" to the exclusion of other categories. The data collected here on study participants' use of gender categories both in self-categorization and in the categorization of others provides support for this hypothesis. Over 99 percent of participants used the categories "male," "female," or a corresponding category.

The second hypothesis was that other gender categories besides "male" and "female" will be used in very small numbers. This hypothesis tested the argument that "male" and "female" are the only categories that a "competent, adult member of society" is morally able to place other into (Frye 1983; Garfinkel 1967; Kessler and McKenna 1978; Ridgeway and Correll 2004). While the participants in this study were younger and more likely to be college educated than US society at large, any usage of gender categories which fall outside the male/female

dichotomy points towards a loosening of the moral requirement of dichotomous gender categories that was predominant in previous decades.

The third hypothesis was that the category "male" would be used disproportionately more frequently than other categories. Due to the gender composition of the stimulus images, participants categorized a set of images where 29.02 percent of images were of male master stimuli. If the "male" category was not used disproportionately more frequently than other categories, then respondent's should have used the category "male" in about 29.02 percent of responses. However, this was not the case for the total number of times the "male" category was used, the total number of times the "male" category was matched with a male master stimuli, or the number of times the "male" category was mismatched with a female master stimuli. This tendency by study participants to over use the "male" category lends support to arguments that the gender categorization process tends to have a bias towards making a male gender categorization (Kessler and McKenna 1978).

The fourth hypothesis was that a gender category would be applied to most, if not all, of the stimulus images presented. This hypothesis tested the argument that gender categorization is the result of a moral or structural imperative, meaning that all members without exception should and must be placed into a gender category (Frye 1983; Garfinkel 1967; Kessler and McKenna 1978; Ridgeway and Correll 2004). According to this argument, the ability to accomplish gender categorization quickly, accurately, and unremarkably is an essential requirement for an individual to be considered a competent, adult member of society (Speer

2005; West and Zimmerman 1987). Members who are unable to accomplish this must account for their failure to categorize gender (Speer 2005).

This hypothesis, and the argument which it tests, was supported by the small number of participants (10.13 percent) who responded so as to avoid making a gender categorization at least once and the small number of total responses where a gender categorization was avoided (2.50 percent). The argument was further supported by participants' accounting for their failure to accomplish gender categorization. Although a formal analysis of the explanations collected in the course of this study are outside the scope of this discussion, a brief reading shows participants in the act of accounting for their non-categorization. In particular, a separate code was created ("Other - No Gender") for non-categorizing responses which accounted for their failure to categorize gender by explaining that it is impossible to determine the gender of an individual by physical appearance alone. It follows, then, that something else is needed to "rightly" determine gender, and if that something else was available, then the respondent would have accomplished their gender categorization.

In answering the question "How are particular parts of the face and the relationship between them used in the accomplishment of gender categorization?", this study sought to test one hypothesis.

The fifth hypothesis was that the mean confidence level and percentage of gender categorizations which match with the master stimuli's self-categorizations would differ depending on which facial image was presented to a participant. Although there was an overall significant correlation between the amount of the face shown in an image and the mean

confidence level and match percentage, three images in particular stood out. First, the mean confidence levels and match percentages for the "mouth - limited" and "mouth - full" images were significantly higher than images which showed about the same amount of face. Second, the mean confidence levels and match percentages for the "jaw line - limited" image were significantly lower than images which showed about the same amount of face.

These results are particularly interesting because the "mouth - full" and "jaw line - limited" images show about the same location on the face - the area around the mouth beneath the nose. However, in the "jaw line - limited" images, the lips are not visible. This points towards the lips and mouth as important visual cues in the gender categorization process. This may also indicate that the visual indicators of gender generally associated with the mouth and lips - "plumpness," size, shape, and presence of lipstick - are socially considered to be more permanent and therefore more reliable indicators of gender than other visual indicators such as eye shape and size (Jackson 1992). The "jaw line - limited" image may have received low ratings relative to similarly sized images because the primary visual indicator of gender in this area - facial hair - is one of the most obviously malleable gender cues available.

The data also indicate that the nose by itself is considered to be a very poor indicator of gender category. This is interesting when compared to Chronicle and colleagues'' (1994) conclusion that the nose in isolation by itself is a very strong indicator of gender.

In answering the questions "How does an individual's membership in other social categories effect their categorizations of others? How does a target's membership in other social categories effect categorizations about them?", this study sought to test one hypothesis.

The sixth hypothesis was that a participant's confidence level and match percentage would vary in relation to their gender, age, and race, and the gender, age, and race of the master stimuli they were categorizing. Due to the nature of the data sample used to test this hypothesis, each of these three variables was coded into binary categories. Participant and stimuli gender were coded as "Male" or "Female," race was coded as "White" or "Non-White," and age was coded as "18 to 26" or "27 and older." Because these variables had to be grouped in a way which combines many disparate groups, more study will have to be undertaken in order to test this hypothesis fully.

The data collected by this study indicates that in the case of all three systems, the category membership of the member performing the categorization tends to have no significant effect on the gender categorization process. However, the gender and race category of the target member being categorized does tend to have a significant effect. Participants in this study reported higher confidence and were more likely to "rightly" categorize a target if that target was male or white.

CHAPTER 9: CONCLUSION

This study advances the fields of sociology and gender studies in three ways. First, it has produced an updated, although not unexpected, description of the categories which members of US society are sorted into on a day-to-day basis. Members still primarily use the categories "male" and "female," but newer categories are also seeing use. Second, this study has added to and enriched the existing body of knowledge about how members of society draw information from visual cues to place one another in social categories. Particularly interesting are the differences in gender categorization which occur using different parts of the lower face, specifically the mouth, lips, chin, and jaw line. Finally, this study has done what very few other studies have done and compared the physical appearance of different gender, age, and racial groups to examine how membership in each group effects membership in the other groups.

One of the goals of feminist research is to find ways in which research can inform practice and vice versa (Kirk and Okazawa-Rey 2010). In examining the methods by which members of US society place one another into gender categories, this study can be read as a proto-manual on how to electively shape one's physical appearance to be categorized as either "female" or "male," or how to mix those characteristics. This information could help both transgender individuals in being perceived as the gender they wish to be perceived as, and queer activists who wish to provide a critique of the binary sex/gender system by presenting an embodied gender that does not easily fall into the categories "female" or "male."

This study will hopefully stand as a starting point for similar studies examining the accomplishment of group membership in day-to-day interaction via visual cues. By broadening the knowledge base about the often overlooked 'categorizing' side of the intersecting social systems which frame day-to-day interaction, this study promotes a kind of fuller understanding of those social systems often necessary in changing them.

APPENDIX A: DEMOGRAPHIC SURVEY QUESTIONS

APPENDIX A: DEMOGRAPHIC SURVEY QUESTIONS

1.1) What is your age? _____

1.2) What is your gender? _____

1.3) What is your race/ethnicity?

(0) American Indian or Alaska Native
(1) Hawaiian or Other Pacific Islander
(2) Asian or Asian American
(3) Black or African American
(4) Hispanic or Latino
(5) Non-Hispanic White
(6) Prefer Not To Respond
(7) Other

1.4) What is the highest level of education you have completed?

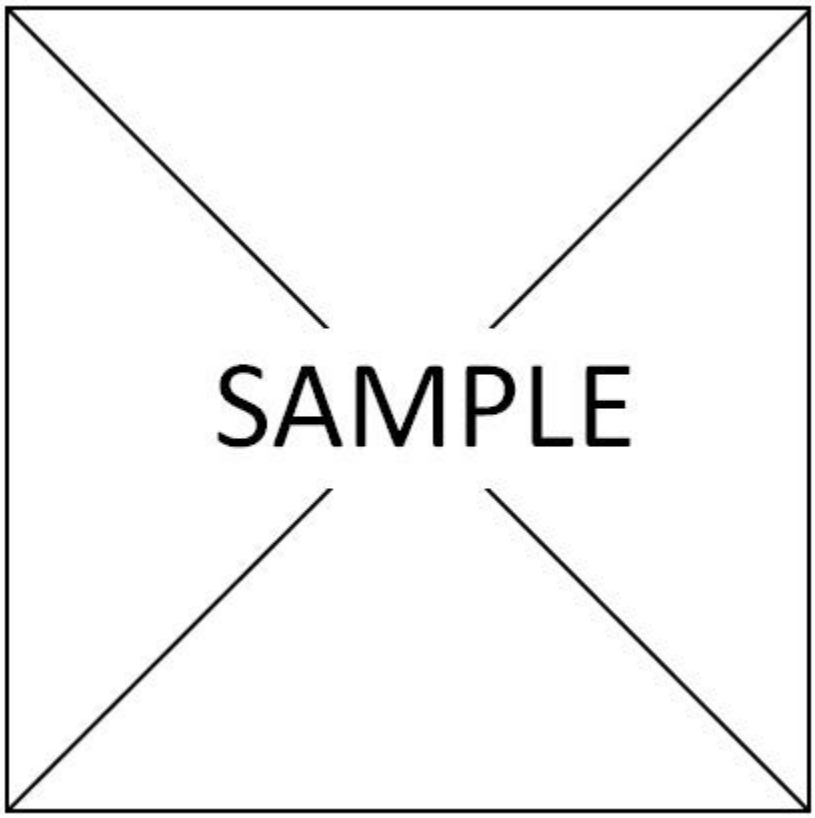
(0) 12^{th} grade or less
(1) Graduated high school of equivalent
(2) Some college, no degree
(3) Associate Degree
(4) Bachelor's Degree
(5) Post-graduate Degree

1.5) If you are currently enrolled in college, what degree program are you enrolled in?

_____

APPENDIX B: STIMULI SURVEY QUESTIONS

APPENDIX B: STIMULI SURVEY QUESTIONS



Question 1: What is the sex/gender of the person in the above image? _____

Question 2: On a scale from 1 to 7, 1 being “Not confident at all” and 7 being “Completely confident”, how confident are you that your answer above is correct?

Not confident -----> -----> -----> -----> Completely confident
1 2 3 4 5 6 7

Question 3: In a sentence or two, please explain the reason for your answers to the two questions above.

_____
_____
_____

APPENDIX C: MASTER STIMULI IMAGES AND STIMULUS IMAGE EXAMPLE

APPENDIX C: MASTER STIMULI IMAGES AND STIMULUS IMAGE EXAMPLE

Table 8: Master Stimuli

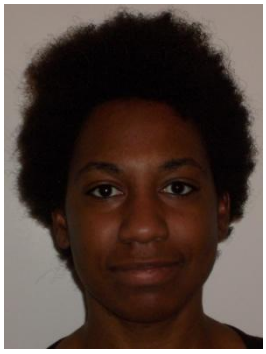
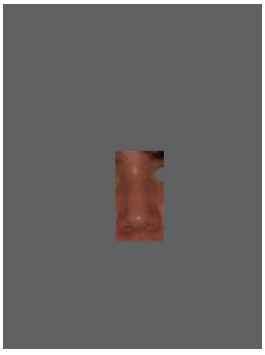
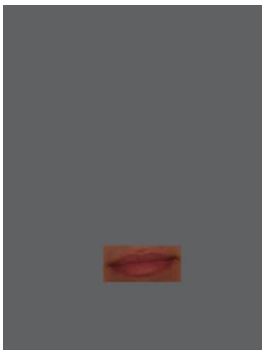
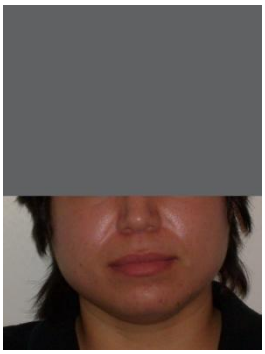
Stimuli 1	Stimuli 2	Stimuli 3	Stimuli 4
			
Stimuli 5	Stimuli 6	Stimuli 7	Stimuli 8
			
	Stimuli 9	Stimuli 10	
			

Table 9: Stimulus Image Examples (Master Stimuli 1)

Facial Oval	Facial Oval - Negative	Brows - Limited
		
Brows - Full	Eyes - Limited	Eyes - Full
		
Nose - Limited	Nose Horizontal - Full	Nose Vertical - Full
		

Mouth - Limited	Mouth - Full	Jaw Line - Limited
		
Lower Face - Full	Upper Face - Full	Full Facial
		

REFERENCES

Barlow, David H., Steven C. Hayes, Rosemery O. Nelson, David L. Steele, Marie E. Meeler, and Joyce R. Mills. 1979. "Sex Role Behavior: A Behavioral Checklist." *Behavioral Assessment.* 1:119-38.

Birdwhistell, Ray L. 1970. "Masculinity and Femininity as Display." Pp. 39-46 in *Kinesics and Context: Essays on Body Motion Communication*, edited by B. Jones. Philadelphia, PA: University of Pennsylvania Press.

Brown, Elizabeth and David I Perrett. 1993. "What Gives A Face Its Gender?" *Perception.* 22:829-40.

Bruce, Vicki, A. Mike Burton, Elias Hanna, Pat Healey, Oli Mason, Anne Coombes, Rick Fright, and Alf Linney. 1993. "Sex Discrimination: How Do We Tell the Difference Between Male and Female Faces?" *Perception.* 22:131-52.

Burton, Mike A., Vicki Bruce, and Neal Dench. 1993. "What's the Difference Between Men and Women? Evidence From Facial Measurement." *Perception.* 22:153-76.

Chronicle, Edward P., Mei-Yin Chan, Charlotte Hawkings, Karen Mason, Kathryn Smethurst, Kate Stellybrass, Kathryn Westrope, and Katie Wright. 1994. "You Can Tell By the Nose - Judging Sex From An Isolated Facial Feature." *Perception.* 24:969-73.

Cotera, Marta. 1997. "Among the Feminists: Racist Classist Issues - 1976." Pp. 213-20 in *Chicana Feminist Thought: The Basic Historical Writings,* edited by A. Garcia. London, England: Routledge.

Crawford, Mary. 2000. “A Reappraisal of Gender: An Ethnomethodological Approach : Editor’s Introduction How to Make Sex and Do Gender.” *Feminism & Psychology.* 10(7):7-10. (Retrieved from Sage Publications on July 29, 2010.)

Crawley, Sara L., Lara J. Foley, and Constance L. Shehan. 2008. *Gendering Bodies.* Lanham, MD: Rowman & Littlefield Publishers, Inc.

Devor, Holly. 1989. *Gender Blending: Confronting the Limits of Duality.* Indianapolis, IN: Indiana University Press.

Dozier, Raine. 2005. “Beards, Breasts, and Bodies: Doing Sex in a Gendered World.” *Gender and Society.* 19(3):297-316. (Retrieved from JSTOR on March 19, 2010.)

Fausto-Sterling, Anne. 2000. *Sexing the Body: Gender Politics and the Construction of Sexuality.* New York: Basic Books.

Frye, Marilyn. 1983. *Politics of Reality: Essays in Feminist Theory.* Trumansburg, NY: Crossing Press.

Garfinkel, Harold. 1967. *Studies in Ethnomethodology.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Goffman, Erving. 1987. *Gender Advertisements.* Harper Torchbooks ed. New York: Harper & Row.

Hausman, Bernice L. 2001. “Recent Transgender Theory.” *Feminist Studies.* 27(2):465-90.

Hawkesworth, Mary. 1997. “Confounding Gender.” *Signs.* 22(3):649-85. (Retrieved from JSTOR on May 25, 2010.)

Heritage, John. 1984. *Garfinkel and Ethnomethodology.* Cambridge, England: Polity Press.

Huart, Johanne, Olivier Corneille, and Emilie Becuart. 2005. "Face-Based Categorization, Context-Based Categorization, and Distortions in the Recollection Of Gender Ambiguous Faces." *Journal of Experimental Social Psychology.* 41:598-608.

Irmen, Lisa. 2006. "Automatic Activation and Use of Gender Subgroups." *Sex Roles.* 55:435-44. (Retrieved from Wilson Web on July 14, 2010.)

Jackson, Linda A. 1992. *Physical Appearance and Gender: Sociobiological and Sociocultural Perspectives.* Albany, NY: State University of New York Press.

Kessler, Suzanne J. 1990. "The Medical Construction of Gender: Case Management of Intersexed Infants." *Signs.* 16(1):3-26.

Kessler, Suzanne J. and Wendy McKenna. 1978. *Gender: An Ethnomethodological Approach.* New York: John Wiley & Sons.

Kirk, Gwen and Margo Okazawa-Rey. 2010. *Women's Lives: Multicultural Perspectives.* 5th ed. New York: McGraw-Hill.

Lang, Sabine. 1999. "Lesbians, Men-Women and Two-Spirits: Homosexuality and Gender in Native American Cultures." Pp. 202-18 in *Constructing Sexualities: Readings in Sexuality, Gender, and Culture,* edited by S. LaFont. Upper Saddle River, NJ: Pearson Education, Inc.

Lorber, Judith. 1991. "The Social Construction of Gender." Pp. 64-7 in *Women's Lives: Multicultural Perspectives,* edited by G. Kirk and M. Okazawa-Rey. New York: McGraw-Hill.

Lorde, Audre. 1980. “Age, Race, Class, and Sex: Women Redefining Difference.” Pp. 454-8 in *Women: Images and Realities, A Multicultural Anthology,* edited by A Kesselman, L. D. McNair, and N. Schniedewind. 4th ed. New York: McGraw-Hill.

Martin, Karin A. 1998. “Becoming a Gendered Body: Practices of Preschools.” *American Sociological Review.* 63(4):494-511. (Retrieved from JSTOR on September 21, 2010.)

McKone, Elinor, Alex Aitkin, and Mark Edwards. 2005. “Categorical and Coordinate Relations in Faces, or Fechner’s Law and Space Instead?” *Journal of Experimental Psychology: Human Perception and Performance.* 31(6):1181-98.

Mealey, Linda. 2000. *Sex Differences: Development and Evolutionary Strategies.* San Diego, CA: Academic Press.

Nanda, Serena. 2000. “Hijra and Sādhim: Neither Man nor Women in India.” Pp. 192-201 in *Constructing Sexualities: Readings in Sexuality, Gender, and Culture,* edited S. LaFont. Upper Saddle River, NJ: Pearson Education, Inc.

Nicholson, Linda J. 1998. “Interpreting ‘Gender’.” Pp. 187-211 in *Race, Class, Gender and Sexuality: The Big Questions,* edited by N. Zack, L. Shrage, and C. Sartwell. Hoboken, NJ: Wiley-Blackwell.

O’Toole, Alice J., Kenneth A Deffenbacher, Dominique Valentine, Karen McKee, David Huff, and Hervé Abdi. 1998. “The Perception of Face Gender: The Role of Stimulus Structure in Recognition and Classificiation.” *Memory and Cognition.* 26(1):146-60.

Oyewumi, Oyeronke. 1998. “De-Confounding Gender: Feminist Theorizing and Western Culture, a Comment on Hawkesworth’s ‘Confounding Gender’.” *Signs.* 23(4):1049-62. (Retrieved from JSTOR on July 29, 2010.)

Quinn, Kimberly A. and C. Neil Macrae. 2005. “Categorizing Others: The Dynamics of Person Construal.” *Journal of Personality and Social Psychology.* 88(3):467-79.

Ridgeway, Cecilia L. 2009. “Framed Before We Know It: How Gender Shapes Social Relations.” *Gender and Society.* 23(2):145-60. (Retrieved from SAGE Journals Online on September 8, 2010.)

Ridgeway, Cecilia L. and Shelley J. Correll. 2004. “Unpacking the Gender System: A Theoretical Perspective on Gender Beliefs and Social Relations.” *Gender and Society.* 18(4):510-31. (Retrieved from JSTOR on March 19, 2010.)

Ridgeway, Ceclilia L. and Lynn Smith-Lovin. 1999. “The Gender System and Interaction.” *Annual Review of Sociology.* 25:191-216. (Retrieved from JSTOR on September 8, 2010.)

Risman, Barbara J. 2004. “Gender as a Social Structure: Theory Wrestling with Activism.” *Gender and Society.* 18(4):429-50. (Retrieved from JSTOR on April 19, 2010.)

Roberts, Tony and Vicki Bruce. 1988. “Feature Saliency in Judging the Sex and Familiarity of Faces.” *Perception.* 17(4):475-81.

Speer, Susan A. 2005. “The Interactional Organization of the Gender Attribution Process.” *Sociology.* 39(1):67-87. (Retrieved from SAGE Journals Online on July 29, 2010.)

Stangor, Charles, Laure Lynch, Changming Duan, and Beth Glass. 1992. "Categorization of Individuals on the Basis of Multiple Social Features." *Journal of Personality and Social Psychology.* 62(2):207-18.

Stokoe, Elizabeth. 2006. "On Ethnomethodology, Feminism, and the Analysis of Categorial Reference to Gender in Talk-In-Interaction." *The Sociological Review.* 54(3):467-94.

West, Candace and Don H. Zimmerman. 1987. "Doing Gender." *Gender and Society.* 1(2):125-51. (Retrieved from JSTOR on May 18, 2010.)

West, Candace and Sarah Fenstermaker. 1995. "Doing Difference." *Gender and Society.* 9(1):8-37. (Retrieved from JSTOR on May 18, 2010.)

Williams, Melissa J. and Gerald A. Mendelsohn. 2008. "Gender Clues and Cues: Online Interactions as Windows into Lay Theories about Men and Women." *Basic and Applied Social Psychology.* 30:278-94.

Yamaguchi, Masami K., Tastu Hirukawa, and So Kanazawa. 1995. "Judgment of Gender Through Facial Parts." *Perception*. 24:563-75.

Unger, Rhoda Kesler. 1979. "Toward a Redefinition of Sex and Gender." *American Psychologist.* 34(11):1085-94.