# Gender Nonconformity and the Stereotype Content Model

Ari M. Rosenblum

Oberlin College

Author Note

Ari M. Rosenblum, Department of Psychology, Oberlin College.

This research was supported in part by the Jerome Davis Research Award and the Oberlin College Department of Psychology.

Abstract

A recent increase in transgender visibility has highlighted gaps in the social psychology literature about attitudes and biases. There is a relatively large body of literature that examines people's reactions to gender role violation, but little that examines reactions to gendered trait violation. To assess negative attitudes towards transgender and gender nonconforming people, this experiment asked participants to make attitude judgements (warmth and competence) about a series of gender stereotypic and counterstereotypic face-voice pairs. This procedure was based on the paradigm used to construct the *Stereotype Content Model*, which categorizes stereotypes/prejudice into four categories (paternalistic, contemptuous, envious, admirable). Participants also rated stimuli on gender using both a continuous (very masculine to very feminine) and categorical (male or female) scale. Overall, counterstereotypic face-voice combinations were rated less warm, but not necessarily less competent than stereotypic face-voice combinations. When plotted in the stereotype content model framework, stimuli that paralleled the gender cues of transfeminine people and non-male-passing transmasculine people were subject to contemptuous and envious prejudice respectively. This kind of prejudice is reflected in the discrimination that transgender people face, including their exclusion from employment and social welfare, as well as their position as a scapegoat for gender-based violence and other socio-political issues.

Gender Nonconformity and the Stereotype Content Model

Most societies have long viewed gender as a biologically predisposed, static characteristic of an individual that fits neatly into the binary categories of “male” and “female.” Across cultures, these binary designations have prescribed behavioral and physical expectations for men and women since before recorded history (Wood & Eagly, 2002). However, as early as the 1950s, stories about Christine Jorgensen, the first widely known American transgender woman to medically transition, began circulating the news and challenging people’s conceptions of gender as an immutable biological fact. Through the 1960s and 1970s, some theorists began to suggest there might be a dissociation between a person’s assigned sex and their innate sense of identity (Diamond, 2004). Although this idea continues to be highly contested, more and more people (academics and the general public alike) are embracing this theory and transgender and gender nonconforming visibility is reaching unprecedented levels (Human Rights Campaign Staff, 2015).

However, a conscious recognition of non-cisgender identities does not necessarily diminish the implicit biases people might have against gender nonconforming people, nor does it prevent them from subconsciously placing gender nonconforming people into existing binary categories. Humans are inherently social beings with an intrinsic ambition to fulfill five core social motives: belonging, understanding, control, esteem, and trust (Fiske, 2013). In order to meet these needs, we have a natural tendency to categorize ourselves and others into distinct social groups. Not only does social categorization fulfill our innate need to belong, but it enhances our perceived ability to predict, understand, and control the outcomes of others’ behavior. Theoretically, group categorizations and stereotypes (i.e., “one’s cognitive expectancies and associations about the group" (Fiske, 2013)) ease social interactions by providing information about another person's goals/motivations (and whether or not they conflict with one's own goals) and/or the person's expected reaction in a given social situation. While stereotypes are often both inaccurate and harmful, people make stereotypical judgements relatively automatically, thus they are extremely difficult to abandon (Fiske, 2013).

But what cues help us make these categorizations so quickly and easily? What happens when an individual violates our expectations of how we believe (consciously or subconsciously) they should act or present themselves? What about when we are exposed to individuals who present ambiguously and we are unable to quickly categorize them within our existing schemas?

Past research on ethnic group categorization suggests that exposure to counterstereotypic members of a group results in less stereotypical evaluations of that group as a whole, implying that, at least to an extent, atypical group members make group categorization more malleable and diminish the salience of the stereotypes associated with their group (Bless, Schwarz, Bodenhausen, & Thiel, 2001). On the other hand, many researchers have found that this effect is moderated by a variety of different factors, including authoritarian inclinations (Adorno, Frenkel-Brunswick, Levinson, & Sanford, 1950), social relevance (Förster, Higgins, & Werth, 2004), self-regulatory focus (Förster et al., 2004), implicit theory of personality (entitativity vs. incrementalism, Plaks, Grant, & Dweck, 2005), and the nature of the stereotype that is being violated (prescriptive vs. descriptive, Rudman & Phelan, 2008). For example, Förster et al. (2004) found that, in conditions of high social relevance, exposure to stereotype disconfirming information about gender roles elicited "agitation-related emotions (i.e., worry and tension)" in participants with a prevention focus and high levels of sexism. Likewise, Rudman and Phelan

(2008) found that violations of *prescriptive* stereotypes (static ideals about how men and women *should* or *should not* be), rather than *descriptive* stereotypes (dynamic perceptions about how men and women *are*) resulted in backlash effects against agentic women in leadership roles. This research ultimately converges on the conclusion that, while moderated by a large number of factors, presenting gender counter-stereotypic representations of individuals to participants is likely to result in discomfort and backlash.

However, while there is a large body of research that examines the violation of stereotypes regarding gender roles and gendered personality traits, there is very little research that examines people's responses to the violation of more physical gendered trait stereotypes, such as facial features, vocal pitch, and prosody (i.e., "the rhythmic and intonational aspect of language" (Merriam-Webster, 2018)). Given the pervasive discrimination that transgender people face in nearly every facet of daily life (especially transgender people who do not "pass" as cisgender, James et al., 2016), it is likely that the backlash effects of disconfirming gender stereotypes about physical characteristics are similar to the effects of disconfirming gender stereotypes about personality traits. To test this hypothesis, this study presented participants with gender stereotypic and counterstereotypic audio-visual stimuli and asked them to categorize the stimuli as either male or female and rate them on masculinity-femininity (continuous gender), warmth, and competence.

This research has the potential to uncover the nature of the stereotypes and prejudice that are directed towards transgender and gender-nonconforming people. Because some of the stimuli in this study parallels many of the gendered cues that transgender women and transfeminine nonbinary people provide (which is not the case for most transmasculine people), this research will be especially applicable to them. These stereotypes and prejudices often result in severe negative outcomes for transgender people, including high rates of poverty, homelessness, sexual and physical violence, murder, HIV infection, substance abuse, mental illness, and suicide (James et al., 2015). In order to combat discrimination and prejudice, it is important to understand how, why, and by whom the stereotypes that inform prejudice are maintained. The results of this study could offer a valuable foundation for future social psychology research addressing prejudice, discrimination, and stereotypes about transgender and gender-nonconforming people.

**Salient Cues in Gender Categorization**

In order to test the research questions, it is important to understand which gendered visual and vocal cues are most salient in the process of gender categorization. There are distinct differences in masculine and feminine facial structures, prosody, and mannerisms that usually offer adequate information to categorize the *perceived* gender of an individual. These variables all fall on a spectrum between two extremes (from “most male” to “most female”), which inevitably makes it more difficult to categorize some people than others. When an individual presents ambiguous gender cues that interfere with our abilities to make automatic categorizations, we tend to look for some kind of auditory or visual trait that offers us salient information. Watson et al. (2012) found that, first and foremost, a person’s voice offers the most salient information for gender categorization. More specifically, researchers have consistently found that the *average pitch* of a person’s voice takes precedence over both visual cues *and* other gendered prosodic cues (Watson et al., 2012). While there is generally an average frequency below which all voices will be categorized as male (~164 Hz) and an average frequency above which all voices will be categorized as female (~181 Hz), there is a range of ambiguous average frequencies that are equally likely to be categorized as both male and female (164-181 Hz). When the average frequency of a voice is within this ambiguous range, people must rely on alternative vocal and/or visual cues to place a person into one of their existing categories (Gelfer & Bennett, 2013). Vocal prosody, which consists of the intonation, resonance, articulation, and vocal quality of a person's voice, often offers salient information about an individual's gender. Intonation refers to the "melodic pattern of an utterance… primarily a matter of variation in the pitch level of the voice" (Britannica, 2017). Feminine voices tend to have a wider pitch range than masculine voices and are characterized by a greater fluctuation in pitch. Resonance refers to the reverberation and fullness of a sound; typically, feminine voices are less resonant or "smaller" than masculine voices (Block, 2017). Articulation refers to the clarity and distinctness of sounds in speech (English Oxford Dictionary, 2017); feminine voices are associated with less articulation than masculine voices. Finally, feminine voices also tend to have a softer and breathier quality than masculine voices (Block, 2017).

Freeman & Ambady (2011) used participants' hand movements ("via the streaming x, y coordinates of the computer mouse") to assess the effect that "sex typical" (e.g. masculine male voices, male faces) and "sex atypical" (e.g. feminine male voices, male faces) face-voice combinations had on binary gender categorizations. They found that, when presented with sex atypical stimuli, participants' hands were initially attracted to the opposite sex category response (i.e. the response that reflected the vocal prosody) before self-correcting to the gender of the face and voice. This suggests that prosody plays a role in perceptions of binary gender when other cues are ambiguous and perceptions of masculinity and femininity, regardless of whether other cues are ambiguous.

**Hypotheses.** Considering the previous research on gender categorization, I expected that faces paired with male-pitched voices would be categorized as male and faces paired with female-pitched voices would be categorized as female, regardless of the gender of the face or the prosody of the voice. However, when the pitch of the voice was in the androgynous range, I expected the gender of the face to be the next most significant predictor of *binary gender* categorization (so long as participants did not perceive the face as androgynous).

Likewise, in determining the *continuous gender* of each face-voice pair, I expected pitch would have the most salient role, followed by prosody, and then facial information. For all face-pitch combinations, I expected feminine prosody to result in more feminine ratings and masculine prosody to result in more masculine ratings (Freeman & Ambady, 2011). Finally, I expected both binary gender categorization and masculinity-femininity ratings to vary based on the extent to which a participant endorsed the gender binary; I predicted that greater endorsement of the gender binary would yield stronger and more significant effects.

## Stereotype Disconfirmation and the Stereotype Content Model

This study used the *Stereotype Content Model* as a framework for examining the backlash effects of disconfirming stereotypes about physical gender characteristics. This also allowed me assess the nature of the stereotypes and prejudice directed towards transgender and gender-nonconforming people. The Stereotype Content Model places stereotypes about social groups on two axes: one that measures perceptions of “warmth” and the other that measures perceptions of “competence.” Groups that are rated low on warmth, but high on competence (e.g.

Asians, Jews, wealthy people, feminists) are subject to “envious stereotypes” and are considered both high status and competitive. Groups that are rated low on competence, but high on warmth (e.g. housewives, elderly people, disabled people), are subject to “paternalistic stereotypes” and are considered both low status and non-competitive. Groups that are rated low on both warmth and competence (e.g. welfare recipients, poor people) are subject to “contemptuous stereotypes” and are considered low status, but competitive. Finally, groups that are rated high on both warmth and competence (e.g. one’s ingroup, close allies) are admired and considered high status and non-competitive (Fiske, Cuddy, Glick, & Xu, 2002). However, no previous research has attempted to place transgender people on this model, which is the goal of this study.

In Rudman and Phelan’s (2008) research, they discussed backlash effects towards both men and women who violated prescriptive gender stereotypes. While women with masculine personality characteristics are rated as less warm than stereotypic women, men who embody more feminine personality characteristics are rated as less competent than stereotypic men. Additionally, boys who exhibit counterstereotypic behavior in childhood face more peer rejection and negative backlash than girls who exhibit counterstereotypic behavior (Martin, 1990).

**Hypotheses.** If the effects of gender role stereotype incongruence match those of gendered trait stereotype incongruence, then Rudman and Phelan’s (2008) research implies that:

1. Male-pitched voices that were paired with feminine prosody should have received lower competence ratings than stereotypically masculine voices.
    a. When these voices were paired with female faces, they should have been rated less warm than when they were paired with male faces. A male-pitched voice with feminine prosody and a female face represented violations of masculine

stereotypes on one hand (feminine prosody and male-pitch) and feminine stereotypes on the other hand (female face and male-pitch).

2. Female-pitched voices that were paired with masculine prosody should have resulted in lower warmth and higher competence ratings than stereotypically feminine voices, especially when paired with male faces.
3. Androgynous-pitched voices with masculine prosody should have been rated more competent than androgynous-pitched voices with feminine prosody, regardless of the gender of the face. Both should have been rated less warm than stereotypic male and female face-voice pairs due to the inability to easily categorize them into a binary gender, but more warm than highly counterstereotypic face-voice pairs because they did not necessarily violate any identifiable prescriptive stereotypes.

Therefore, I predicted that counterstereotypic pairs with male-pitched voices would be subject to more contemptuous prejudice than stereotypic pairs with male-pitched voices, and counterstereotypic pairs with female-pitched voices would be subject to more envious prejudice than stereotypic pairs with female-pitched voices.

# Methods

## Participants

Participants were recruited through the crowdsourcing website *MTurk* and were paid $3.00 for their participation. Based on a power analysis calculated using a similar study that yielded effect sizes (*Cohen's d)* from 0.3 to 1.1 on various dependent variables ( Freeman & Ambady, 2011), a sample size of at least 71 people was needed to minimize the likelihood of making a Type II error. In total, 97 *MTurk* recruits participated in the experiment. All participants signed a consent document and were informed that they may quit the study at any time without deduction from their compensation. Of the 97 participants, 59.8% were male and 40.2% were female. There were no self-identified transgender or nonbinary people who participated in the experimental study. 91.8% of the sample identified as heterosexual, compared to 8.2% who identified as "homosexual, polysexual, or something else." 85.6% of participants indicated that they were white, compared to 14.4% not white. The average political orientation on a 7-point likert scale (1 = *extremely conservative*, 7 = *extremely liberal*) was 4.91 (SD = 1.733); the distribution of political orientation was significantly negatively skewed with a skewness of -.676 (*SE* = .245). Financial status was normally distributed with an average of 3.62 (SD = 1.318) on a 7-point Likert scale (1 = *economically disadvantaged*, 7 = *economically well-off*).

**Procedure**

**Stimuli creation.** The stimuli were created for a 2 (facial cue) x 3 (pitch) x 2 (prosody) within-subjects design. I created the face stimuli by morphing four randomly selected faces together in the computer program *Fantamorph*. In order to control for potential effects of race on participants' responses, all the faces used to create the composite face stimuli were white. There were two categories of composite faces: male (M) and female (F). The composite faces were comprised of either 4 randomly selected male faces or 4 randomly selected female faces from the *Radboud Face Database*. I created six composite faces for each of the categories, totalling to 12 different composite faces *(Appendix A*). Individual faces were not removed from the face bank after each composite face was created, therefore some faces were used to create more than one composite face.

I created the voice stimuli by recording exemplars with stereotypically masculine and feminine vocal prosody saying medium length neutral sentences and using the computer program *Praat* to manipulate the pitch of each voice to the average male ($X_M$), female ($X_F$), and androgynous ($X_A$) frequency. Therefore, each of the 3 pitch categories contained a voice with more masculine prosody ($X_{Mm}$, $X_{Am}$, $X_{Fm}$) and a voice with more feminine prosody ($X_{Mf}$, $X_{Af}$, $X_{Ff}$), totalling to 6 voice/prosody combinations (Note: $X$ represents face gender [M, F], uppercase subscripts represent pitch, and lowercase subscripts represent prosody).

To recruit exemplars with stereotypically masculine and feminine voices, I recorded students from two Psychology classes at Oberlin College saying a medium length sentence. I adjusted the average frequency of each recording to 170 Hz to ensure that vocal pitch did not affect perceptions of masculinity or femininity. My faculty advisor and I listened to the recordings and selected voices that provided the best example of masculine and feminine prosody. Voice samples with a discernible accent were discarded. We then recruited the people whose voices we selected to record additional sentences. We recorded each voice actor saying 24-different neutral sentences with an average length of 22 words (*Appendix B*). To ensure that the sentences used for the voice stimuli were neutral, I ran them through the computerized text analysis software *Linguistic Inquiry and Word Count* (LIWC) and analyzed for emotional valence (either 0 positively and 0 negatively valenced words *or* an equal number of positively and negatively valenced words). All voice actors provided consent to participate in the study and received $10.00 as compensation for their participation.

**Pilot testing methods.** All stimuli were subject to pilot testing by naïve participants to ensure reliability. Pilot testing participants were recruited through *Amazon Mechanical Turk*

(*MTurk*) and completed the Pilot testing survey on *Qualtrics*. Initial screening ensured that all participants lived in the United States, were 18 years or older, fluent in English, and had normal or corrected to normal vision and hearing. All participants completed the Pilot test on a computer (rather than a mobile device) either wearing headphones or out loud in a quiet, low distraction environment. Pilot testing participants received compensation of $2.00 for their participation. Participants were reminded that their participation in the survey was entirely voluntary and that they could skip any questions that they did not wish to answer or stop the survey at any time and still receive compensation. Participants completed the following tasks and then answered a series of demographic questions.

*Task 1: Face Binary and Continuous Gender Ratings*

In *Task 1* participants were shown 12 faces (6 male, 6 female) in a random order and were then asked to rate the faces on binary and continuous gender. Participants received the following instructions:

> *For this first task, you will be shown a series of faces. Below each face, you will be asked to rate the face on two different dimensions. Please indicate your judgments about each face as quickly as possible.*

Participants then answered the following questions about each face:

1. *What gender is this individual?* (*Male* or *Female*)
2. *How masculine or feminine is this individual?* (1 = *very masculine*, 5 = *very feminine*)

*Task 2: Voice Binary and Continuous Gender Ratings*

In *Task 2*, participants heard 2 randomly selected audio recordings from each of the 9 voice actors (5 masculine, 4 feminine) recorded for the study (presented in a random order). The pitch of each voice recording in the survey was manipulated to have a median frequency of 165 Hz. They were then asked to rate these voices on binary and continuous gender. This task was run 3 separate times, each time with clearer instructions and changes to the stimuli and/or rating scale. The first time this task was run, recordings of male and female voice actors had been manipulated to have a median pitch of 180 Hz and 165 Hz respectively. The difference in median pitch was intended to limit participants' ability to use pitch (as opposed to prosody) to judge voices; however, results from the first trial suggested that this did the exact opposite. During the first trial of this task, participants received the following instructions:

> *For this second task, you will listen to a series of audio recordings. Press play and please only listen to each audio file once. After you have listened to each audio file, please rate each voice on the two dimensions below. Please indicate your judgments about each voice as quickly as possible.*

Participants then answered the following questions about each voice:

1. *What gender is this voice?* (*Male* or *Female*)
2. *How masculine or feminine is this voice?* (1 = *very masculine*, 5 = *very feminine*)

For the second time this task was run, the pitch of all voice was manipulated to have the same median frequency (165 Hz) and participants received the following instructions:

> *For this second task, you will listen to a series of audio recordings. Press play and please only listen to each audio file once. After you have listened to each audio file, please rate*

*each voice on the two dimensions below. When you consider your judgments, try to pay attention to the vocal inflections of the voice and **not** the pitch of the voice. Please indicate your judgments about each voice as quickly as possible.*

Participants then answered the following questions about each voice:

1. *What gender is this voice?* (*Male* or *Female*)
2. *How masculine or feminine is this voice?* (1 = *very masculine*, 5 = *very feminine*)

The third time this task was run, participants rated each voice's continuous gender on a 7-point Likert scale instead of a 5-point Likert scale. Past studies have shown that white Americans (who comprise the largest demographic group in our sample) are unlikely to pick the "extremes" on a 5-point Likert scale (i.e. 1 or 5) (Hui & Triandis, 1989). This problem appeared to have affected the pilot data and did not allow for much of any variation in the average ratings for each voice, which is why the third trial used a 7-point scale. Participants also received the following, more detailed instructions:

*For this second task, you will listen to a series of audio recordings. Press play and please only listen to each audio file once. After you have listened to each audio file, please rate each voice on the two dimensions below. The voices you will hear are actual recordings of men's and women's voices. The pitch of all voices has been transposed to be in the same frequency range. When you consider your judgments, try to pay attention to the vocal inflections of the voice and **not** the pitch of the voice. Please indicate your judgments about each voice as quickly as possible.*

Participants then answered the following questions about each voice:

1. *What gender is this voice?* (*Male* or *Female*)

2. *How masculine or feminine is this voice?* (1 = *very masculine*, 7 = *very feminine*)

I only used the data from the third trial of *Task 2* to assess the reliability of the stimuli. This task intended to verify that the masculine voices were perceived as distinctly masculine and feminine voices as distinctly feminine.

*Demographic Information and Debriefing*

After completing the tasks, participants were asked to answer a series of demographic questions (see *Appendix C*). Participants were then debriefed about the intentions of the study and reminded that they could remove their data from dataset until March 15, 2018 without penalty to their compensation.

**Pilot testing results.** A total of 90 participants completed Task 1. Of those participants, 72% were white (28% non-white). 60% of participants were male, 39% female, and 1% non-binary. Political orientation was significantly negatively skewed (*M* = 4.7, *SD* = 1.675) (1 = *extremely conservative*, 7 = *extremely liberal*). Only 19% of participants considered themselves moderately to extremely conservative, compared to 24.7% of participants who indicated that they were politically neutral and 56.3% who indicated that they were moderately to extremely liberal. 95.5% of respondents indicated that they identified as heterosexual (4.5% non-heterosexual). A total of 30 participants completed *Task 2*. Of those participants, 70% were white (30% non-white), 63.3% male (36.7% female), and 100% identified as heterosexual. 20% of respondents indicated that they were politically neutral, while 26.6% indicated that they were moderately to extremely conservative and 53.3% indicated that they were moderately to extremely liberal (*M* = 4.63, *SD* = 1.691).

*Task 1: Face Stimuli Gender*

Participants categorized all faces as their intended binary gender at least 98.9% of the time. *Figure 1* shows the average continuous gender rating for each face with a 95% confidence interval around each mean. As seen in *Figure 1*, every female face was rated significantly more feminine than every male face. Thus, all face stimuli were reliably categorized as the gender they intended to represent.

*Task 2: Voice Stimuli Gender*

A Goodness of Fit test revealed that most of the voices were significantly more likely to be rated their target gender than the opposite gender, $\chi^2(1) \geq 7.759$ , $p \leq .005$. Binary gender categorizations for two voices were not significantly different from what we would expect by chance, $\chi^2(1) \leq .862$, $p > .05$. One voice was rated the opposite of its target gender marginally more frequently that its target gender ($p = .059$), while another was rated the opposite of its target gender significantly more frequently than its target gender ($\chi^2(1) = 4.172$ , $p < .05$)(See *Table 1*).

Additionally, *Figure 2* shows the average continuous gender rating for each voice with a 95% confidence interval, as well as the overall mean, the masculine voice mean, and the feminine voice mean. Three voices were neither significantly different from all opposite gender voices nor significantly different from the overall mean. In order to ensure that all the voice stimuli had distinctly masculine or feminine prosody, these voices were removed from the stimulus pool.

In sum, pilot testing identified 12 faces (6 male, 6 female) and 6 voices (3 masculine, 3 feminine) that were reliably identified as intended. In order to select stimuli for each of the possible face-voice combinations shown in *Table 2*, a face and voice were randomly selected from their respective categories (i.e., for $F_{Mm}$, a female face and a male-pitched masculine voice were randomly selected and paired). Without replacement, a second face and voice were randomly selected from the same categories to create a second face-voice pair for that combination. In total, there were 24 face-voice pairs, consisting of 2 face-voice pairs for each combination in *Table 2*. The neutral sentence used for each voice was randomly assigned among the 24 face-voice pairs without repetition (participants heard each sentence only once per task).

**Main experiment.** The experimental survey was administered using *Qualtrics*. In order to be eligible to complete the survey, all participants were required to be 18 years or older and give their consent to participate. Additionally, participants were asked to confirm that they were fluent English speakers, had normal or corrected to normal vision and hearing, that they were completing the task on a computer, and that they were either using headphones or were in a quiet environment. Participants who did not meet these requirements were excluded from participation and were not compensated for their time.

Participants rated all 24 face-voice pairs in a random order. When a participant clicked the “next” button, the face appeared and participants were prompted to press play and only listen to the audio recording once. Participants answered 2 separate sets of questions for each of these face-voice pairs. The first set of questions assessed the participant’s perceptions of warmth and competence, while the second set assessed perceptions of categorical and continuous gender.

For the first task, participants read the following instructions and answered the following questions about each stimuli presented:

*For this first task, you will see and hear a series of face-voice pairs. Press play and please only listen to each audio file once. Try to imagine that the voice you hear is that of the person depicted with it, even if some of the combinations may seem strange. After you have listened to the audio file, please rate the face-voice pair on the two dimensions below. Please indicate your judgments about each pair as quickly as possible.*

1. *How competent do you think this person is?* (1= *not at all*, 5= *extremely*) (Fiske, Cuddy, Glick, & Xu, 2002)
2. *How warm do you think this person is?* (1= *not at all*, 5= *extremely*) (Fiske et al., 2002)

For the second task, participants read the following instructions and answered the same questions as *Task 1* in the pilot test.

*For this second task, you will again see and hear a series of face-voice pairs. Press play and please only listen to each audio file once. Try to imagine that the voice you hear is that of the person depicted with it, even if some of the combinations may seem strange. After you have listened to the audio file, please rate the face-voice pair on the two different dimensions below. Please indicate your judgments about each pair as quickly as possible.*

Participants were informed that they could skip any questions that they did not wish to answer. After completing the ratings for each of the face-voice pairs presented to them, *MTurk* participants were asked to complete a post-screen questionnaire. The questionnaire included a variety of demographic questions (age, race, gender, socioeconomic status, etc.) as well as questions pertaining to political ideology, endorsement of the gender binary, authoritarian tendencies, implicit theory of personality, regulatory focus, transphobic beliefs, tolerance/intolerance of ambiguity, and exposure to transgender and gender nonconforming people. The post-screen questionnaire and scale reliability reports (where relevant) can be found in *Appendix E*. After the questionnaire participants were debriefed about the intentions of the study and thanked for their participation.

## Results

Analyses of the post-screen attitude measures revealed that, contrary to my expectations, transphobia, authoritarianism, tolerance-intolerance of ambiguity, regulatory focus, implicit person theory, and endorsement of the gender binary did not have significant effects on binary gender categorization, continuous gender, warmth, or competence ratings. Demographic factors (i.e. race, gender, political orientation, financial status, and sexual orientation) also did not have significant effects on any of the dependent variables.

### Binary Gender

In order to determine which stimulus combinations were more likely to be categorized as male or female, I coded *male* categorizations as 0, and *female* categorizations as 1. I then computed average binary gender variables for each of the stimulus combinations. I did this by averaging the binary gender scores for the two versions of each stimulus category into one variable. For example, the binary gender score for $F_{Ff\,1}$ and $F_{Ff\,2}$ were averaged to make the variable $F_{Ff}$_ BinaryGender. These new variables represent the percent of the time that a stimulus combination was categorized as female. A combination with a mean binary gender score of 0 was categorized as male 100% of the time, while a combination with a mean binary gender score of 1 was categorized as female 100% of the time. To assess which face-voice combinations were more likely to be categorized as male or female, I ran one-sample t-tests for each stimulus combination, comparing it to an expected mean of .5. This expected mean represents the null hypothesis that a face-voice combination is equally likely to be categorized as male and female. Except for $F_{Mm}$ (female face, male-pitch, masculine prosody), all stimuli combinations with a female face were significantly more likely to be categorized as female than male. Likewise, except for $M_{Ff}$ (male face, female-pitch, female prosody), all stimuli combinations with a male face were significantly more likely to be categorized as male than female (See *Table 3* for descriptives and *Table 4* for test statistics and significance values). To determine the impact of gender categorization on continuous gender, warmth, and competence ratings, I averaged all continuous gender scores together, warmth scores together, and competence scores together for each gender category (Male, Female, Androgynous). This created continuous gender, warmth, and competence scores for male categorized stimuli, female categorized stimuli, and androgynous categorized stimuli ($F_{Mm}$, $M_{Ff}$).

**Continuous Gender**

I ran a 2 x 3 x 2 (face (2), pitch (3), prosody (2)) repeated measures analysis of variance (ANOVA) to determine whether face, pitch, and prosody had an effect on continuous gender ratings. I found significant main effects for pitch and prosody, but not for face. Participants rated masculine prosody voices significantly more masculine than feminine prosody voices ($M_m$ = 3.337, $SD_m$ = .339, $M_f$ = 2.070, $SD_f$ = .404). Paired sample t-tests revealed that participants rated male pitched voices ($M_M$ = 3.445, $SD_M$ = .406) significantly more masculine than both female-pitched ($M_F$ = 1.991, $SD_F$ = .390) ($t(96) = 23.116$, $p < .001$, $d = 3.654$) and androgynous pitched voices ($M_A$ = 2.675, $SD_A$ = .345) ($t(96) = 16.971$, $p < .001$, $d = 2.042$), and that participants rated androgynous pitched voices significantly more masculine than female pitched voices ($t(96) = 16.255$, $p < .001$, $d = 1.860$). These are all very large effects, as there is greater than a 98.31% chance that a male-pitched voice chosen at random was rated more masculine than a female-pitched voice chosen at random, a 92.14% chance that a male-pitched voice chosen at random was rated more masculine than an androgynous pitched voice chosen at random, and a 90% chance that an androgynous-pitched voice chosen at random was rated more masculine than a female-pitched voice chosen at random (probability of superiority). Overall, pitch and prosody all had large effects on continuous gender ratings. Pitch accounted for 80.9% of the variance remaining after excluding the proportion of variance due to other effects, and prosody accounted for 83.9% of the variance remaining after excluding the proportion of variance due to other effects.

Additionally, there were significant interactions between face and pitch, face and prosody, pitch and prosody, and face, pitch, and prosody. The face by pitch interaction had a relatively small effect, accounting for only 4.4% of the variance in continuous gender ratings after excluding the proportion of variance due to other effects. The face by prosody interaction had a moderate effect, accounting for only 8.7% of the variance in continuous gender ratings after excluding the proportion of variance due to other effects. The pitch by prosody interaction had a moderate effect, accounting for 6.2% of the variance in continuous gender ratings after excluding the proportion of variance due to other factors. The face by pitch by prosody interaction had a large effect, accounting for 14.9% of the variance in continuous gender ratings after excluding the proportion of variance due to other factors. See *Table 5* for ANOVA test-statistics. The interactions between face and pitch, face and prosody, and pitch and prosody had small effects that were unrelated to my hypotheses. See *Appendix D* for post-hoc analyses of these interactions.

Post-hoc paired sample t-tests for the interaction between face, pitch, and prosody revealed that the gender of the face only played a significant role in continuous gender ratings for female-pitched voices (both masculine and feminine prosody). Stimuli with female-pitched voices, feminine prosody, and male faces ($M_{Ff}$) were rated significantly more masculine than stimuli with female-pitched voices, feminine prosody, and a female faces ($F_{Ff}$) ($t(96) = 4.012$, $p < .001$, $d = .528$). For stimuli with female-pitched voices and masculine prosody, female faces ($F_{Fm}$) were rated significantly more masculine than male faces ($M_{Fm}$) ($t(96) = 37.857$, $p < .001$, $d = .521$). Androgynous-pitched voices with both masculine ($F_{Am}$, $M_{Am}$) and feminine prosody ($F_{Af}$, $M_{Af}$) were rated equally as masculine for both male and female faces ($t_m(96) = 1.355$, $p > .05$, $d_m = .216$) ($t_f(96) = 1.367$, $p > .05$, $d_f = .192$). Likewise, male-pitched voices with both masculine ($F_{Mm}$, $M_{Mm}$) and feminine prosody ($F_{Mf}$, $M_{Mf}$) were rated equally masculine for both male and female faces ($t_m(96) = -1.674$, $p > .05$, $d_m = .207$) ($t_f(96) = .353$, $p > .05$, $d_f = .050$). For graphs of this interaction, see *Figure 6*. For descriptive statistics, see *Table 6*.

I also ran repeated measures ANOVAs to determine whether the simplified female ($M_F = 2.454$, $SD_F = .510$), androgynous ($M_A = 2.830$, $SD_A = .399$), and male ($M_M = 2.909$, $SD_M = .425$) categorized stimuli pairs differed on continuous gender. There was a significant main effect of Gender category ($F(2) = 27.444$, $p < .001$, $\eta^2 = .222$). This was a large effect, as it accounted for 22.2% of the variance in continuous gender ratings after excluding the proportion of variance due to other effects. Post-hoc paired sample t-tests revealed that both male categorized and androgynous categorized stimuli were rated more masculine than female categorized stimuli ($t(96) = 5.846$, $p < .001$, $d = .969$) ($t(96) = 6.483$, $p < .001$, $d = .821$). Male categorized and androgynous categorized stimuli were rated equally as masculine ($t(96) = 1.317$, $p > .05$, $d = .192$).

**Warmth**

I ran a 2 x 3 x 2 (face (2), pitch (3), prosody (2)) repeated measures ANOVA to determine whether face, pitch, and prosody had an effect on warmth ratings. I found significant main effects for face, pitch, and prosody. Participants rated feminine prosody voices significantly more warm than masculine prosody voices ($M_f = 3.199$, $SD_f = .436$, $M_m = 2.790$, $SD_m = .466$), and male faces as significantly more warm than female faces ($M_M = 3.052$, $SD_M = .424$, $M_F = 2.938$, $SD_F = .445$). Paired sample t-tests revealed that female-pitched voices ($M_F = 3.151$, $SD_F = .461$) were rated significantly more warm than both male-pitched ($M_M = 2.990$, $SD_M = .479$) ($t(96) = 3.266$, $p < .01$, $d = .342$) and androgynous-pitched voices ($M_A = 2.844$, $SD_A = .510$) ($t(96) = 6.022$, $p < .01$, $d = .630$), and that male-pitched voices were rated significantly more warm than androgynous-pitched voices ($t(96) = 3.431$, $p < .01$, $d = .294$). The effect of pitch on warmth ratings for female-pitched vs. male-pitched voices and male-pitched voices vs. androgynous-pitched voices was relatively small, as there is only a 59% chance that a female-pitched voice chosen at random was rated more warm than a male-pitched voice chosen at random, and a 58% chance that a male-pitched voice chosen at random was rated more warm than an androgynous pitched voice chosen at random (probability of superiority). However, the effect of pitch on warmth ratings for female-pitched voices vs. androgynous pitched voices was moderate; there is a 67% chance that a female-pitched voice chosen at random was rated more warm than an androgynous-pitched voice chosen at random (probability of superiority). Overall, pitch and prosody had large effects on warmth ratings. Pitch accounted for 17.7% of the variance remaining after excluding the proportion of variance due to other effects and prosody accounted for 49.6% of the variance remaining after excluding the proportion of variance due to other effects. Face had a moderate effect on warmth ratings, accounting for 10.1% of the variance remaining after excluding the proportion of variance due to other effects.

In addition to the main effects, significant interactions were found between face and pitch, face and prosody, pitch and prosody, and face, pitch, and prosody. The face by pitch, face by prosody, and pitch by prosody interactions had large effects. The face by pitch interaction accounted for 13.9% of the variance in warmth ratings after excluding the proportion of variance due to other effects. The face by prosody interaction accounted for 60.2% of the variance remaining after excluding the proportion of variance due to other effects. The pitch by prosody interaction accounted for 53.2% of the variance in warmth ratings after excluding the proportion of variance due to other effects. The face by pitch by prosody interaction had a moderate effect, accounting for 9.4% of the variance in warmth ratings after excluding the proportion of variance due to other effects. See *Table 7* for ANOVA test-statistics

Post-hoc paired sample t-tests for the interaction between face and pitch revealed that female faces with female-pitched voices ($M = 3.211$, $SD = .548$) were rated more warm than male faces with female-pitched voices ($M = 3.090$, $SD = .521$) ($t(96) = 2.2$, $p < .05$, $d = .226$), while female faces with androgynous-pitched voices ($M = 2.673$, $SD = .602$) were rated significantly less warm than male faces with androgynous-pitched voices ($M = 3.016$, $SD = .574$) ($t(96) = 5.814$, $p < .001$, $d = .583$). Male faces with male-pitched voices ($M = 3.049$, $SD = .529$)

were also rated marginally more warm than female faces with male-pitched voices (*M* = 2.930, *SD* = .611) ($t(96) = -1.889$, $p < .1$, $d = .208$). This suggests that stimuli with stereotype congruent face and pitch were considered more warm than stimuli with incongruent face and pitch. For a graph of this interaction, see *Figure 7*.

Post-hoc paired sample t-tests for the interaction between face and prosody revealed that the gender of the face only effected warmth ratings for feminine prosody voices. Stimuli with feminine prosody voices were rated significantly more warm when paired with male faces (*M* = 3.287, *SD* = .530) than when paired with female faces (*M* = 3.112, *SD* = .493) ($t(96) = 3.182$, $p < .01$, $d = .342$). Stimuli with masculine prosody voices were rated equally as warm, regardless of face gender ($M_{MXm} = 2.816$, $SD_{MXm} = .485$, $M_{FXm} = 2.765$, $SD_{FXm} = .553$) ($t(96) = 1.106$, $p > .05$, $d = .098$). Therefore, while people consider male faces with female-pitched voices to be less warm than stereotype congruent stimuli, incongruent prosody did not result in lower warmth ratings. For a graph of this interaction, see *Figure 8*.

Post-hoc paired sample t-tests for the interaction between pitch and prosody revealed that prosody significantly impacted warmth ratings for male-, androgynous-, and female-pitched voices, but in different directions. For female- and androgynous-pitched voices, feminine prosody voices ($M_{XFf} = 3.706$, $SD_{XFf} = .607$, $M_{XAf} = 3.008$, $SD_{XAf} = .603$) were rated significantly more warm than masculine prosody voices ($M_{XFm} = 2.595$, $SD_{XFm} = .578$, $M_{XAm} = 2.680$, $SD_{XAm} = .587$) ($t_F(96) = 14.618$, $p < .001$, $d_F = 1.875$) ($t_{XA}(96) = 5.274$, $p < .001$, $d_A = .551$). For male-pitched voices, however, masculine prosody voices ($M_{XMm} = 3.095$, $SD_{XMm} = .558$) were rated significantly more warm than feminine prosody voices ($M_{XMf} = 2.884$, $SD_{XMf} = .576$) ($t(96) = 3.403$, $p < .01$, $d = .372$). The combination of feminine pitch and feminine prosody was rated particularly high on warmth. This interaction again suggests that stereotype congruent stimuli were considered more warm than stereotype incongruent stimuli. For a graph of this interaction, see *Figure 9*.

I also ran post-hoc paired sample t-tests for the interaction between face, pitch, and prosody. Female-pitched voices with feminine prosody ($F_{Ff}$, $M_{Ff}$) were rated equally as warm, regardless of the gender of the face ($t(96) = 1.226$, $p > .05$, $d = .140$), while androgynous-pitched voices and male-pitched voices with feminine prosody were considered significantly more warm when paired with male faces than female faces ($t_A(96) = -2.846$, $p < .01$, $d_A = .311$) ($t_M(96) = 2.072$, $p < .05$, $d_M = .269$). When female-pitched voices were paired with masculine prosody, female faces were rated significantly more warm than male faces ($t(96) = 4.726$, $p < .001$, $d = .508$). On the other hand, androgynous-pitched voices with masculine and feminine prosody were rated more warm when paired with male faces ($M_{Am}$, $M_{Af}$) than with female faces ($F_{Am}$, $F_{Af}$) ($t_m(96) = 5.949$, $p < .001$, $d_m = .660$) ($t_f(96) = 2.846$, $p < .01$, $d_f = .311$). Stimuli with male-pitched voices and masculine prosody ($M_{Mm}$, $F_{Mm}$) were rated equally warm, regardless of whether they were paired with male or female faces ($t(96) = .409$, $p > .05$, $d = .051$). To summarize, stimuli with female-pitched feminine voices ($X_{Ff}$) and male-pitched masculine voices ($X_{Mm}$) were considered relatively high on warmth regardless of face, implying that participants preferred stimuli with matching pitch and prosody over stimuli with mismatching pitch and prosody. Other than female-pitched masculine voices with male faces, which were rated very low on warmth, all other pitch-prosody combinations were rated more warm when paired with male faces than female faces. For graphs of this interaction, see *Figure 10*. For descriptive statistics, see *Table 8*.

In order to determine whether the stimuli that participants categorized as female, androgynous, and male differed on warmth ratings, I ran another repeated measures ANOVA with only one factor (gender categorization). There was a significant main effect of Gender category ($F(2) = 65.585$, $p < .001$, $\eta^2 = .406$). This was a large effect, as it accounted for 40.6% of the variance in warmth ratings after excluding the proportion of variance due to other effects. Post-hoc paired sample t-tests revealed that androgynous categorized stimuli ($M_A = 3.418$, $SD_A = .616$) were rated significantly more warm than female ($M_F = 2.907$, $SD_F = .481$) ($t(96) = 8.793$, $p < .001$, $d = .925$), and male ($M_M = 2.910$, $SD_M = .435$) categorized stimuli ($t(96) = 9.218$, $p < .001$, $d = .953$). Male categorized and female categorized stimuli were rated equally warm ($t(96) = .077$, $p > .05$, $d = .007$).

**Competence**

I ran a 2 x 3 x 2 (face (2), pitch (3), prosody (2)) repeated measures ANOVA to determine whether face, pitch, and prosody had an effect on competence ratings. I found significant main effects for face, pitch, and prosody. Male faces were rated more competent than female faces ($M_M = 3.454$, $SD_M = .595$ , $M_F = 3.377$, $SD_F = .551$) and feminine prosody voices were rated significantly more competent than masculine prosody voices ($M_f = 3.475$, $SD_f = .553$, $M_m = 3.356$, $SD_m = .625$). Post-hoc paired sample t-tests revealed that male-pitched voices ($M_M = 3.483$, $SD_M = .586$) were rated significantly more competent than female-pitched voices ($M_F = 3.214$, $SD_F = .596$) ($t(96) = 6.010$, $p < .001$, $d = .456$), and androgynous-pitched voices ($M_A = 3.549$, $SD_A = .631$) were rated significantly more competent than female pitched voices ($t(96) = 6.472$, $p < .001$, $d = .546$). Male-pitched voices and androgynous-pitched voices were rated equally as competent ($t(96) = 1.782$, $p > .05$, $d = .108$). The effect of pitch on competence ratings for female-pitched vs. male-pitched voices and female-pitched voices vs. androgynous-pitched voices was relatively moderate. There is a 62% chance that a male-pitched voice chosen at random was rated more competent than a female-pitched voice chosen at random and a 65% chance that an androgynous-pitched voice chosen at random was rated more competent than a female-pitched voice chosen at random (probability of superiority). Overall, face and prosody had small to moderate effects on competence ratings. Face accounted for 5% of the variance remaining after excluding the proportion of variance due to other effects. Prosody accounted for 7% of the variance remaining after excluding the proportion of variance due to other effects. Pitch had a large effect on competence ratings, accounting for 24.6% of the variance remaining after excluding the proportion of variance due to other effects.

In addition to the main effects, significant interactions were found between face and pitch, pitch and prosody, and face, pitch, and prosody. The interaction between face and prosody was not significant. The face by pitch interaction had a moderate effect on competence ratings, accounting for 10.5% of the variance remaining after excluding the proportion of variance due to other effects. The pitch by prosody interaction and the face by pitch by prosody interaction both had large effects on competence ratings. The pitch by prosody interaction accounted for 25.1% of the variance remaining after excluding the proportion of variance due to other effects, while the face by pitch by prosody interaction accounted for 23.7% of the variance remaining after excluding the proportion of variance due to other effects. See *Table 7* for ANOVA test-statistics.

Post-hoc paired sample t-tests for the interaction between pitch and face on competence ratings revealed that the gender of the face only impacted competence ratings for androgynous pitched voices. Stimuli with female-pitched voices were rated equally as competent when paired with both male ($M_{MFx}$ = 3.170, $SD_{MFx}$ = .669)and female faces ($M_{FFx}$ = 3.258, $SD_{FFx}$ = .642)($t$(96) = 1.571, $p$ > .05, $d$ = .134). Likewise, stimuli with male-pitched voices were rated equally as competent when paired with both male ($M_{MMx}$ = 3.490, $SD_{MMx}$ = .645) and female faces ($M_{FMx}$ = 3.477, $SD_{FMx}$ = .656) ($t$(96) = .228, $p$ > .05, $d$ = .020). However, face-voice pairs with androgynous-pitched voices were rated significantly more competent when paired with a male face ($M_{MAx}$ = 3.701, $SD_{MAx}$ = .730) than when paired with a female face ($M_{FAx}$ = 3.397, $SD_{FAx}$ = .694) ($t$(96) = 4.537, $p$ < .001, $d$ = .427). For a graph of this interaction, see *Figure 11.*

I also ran post-hoc paired sample t-tests for the interaction between pitch and prosody. Stimuli with female-pitched voices and feminine prosody ($M_{XFf}$ = 3.446, $SD_{XFf}$ = .675) were rated significantly more competent than stimuli with female-pitched voices and masculine prosody ($M_{XFm}$ = 2.982, $SD_{XFm}$ = .734) ($t$(96) = 2.072, $p$ < .05, $d$ = .658). Androgynous-pitched voices were rated equally as competent for both masculine ($M_{XAm}$ = 3.516, $SD_{XAm}$ = .764) and feminine prosody voices ($M_{XAf}$ = 3.583, $SD_{XAf}$ = .644) ($t$(96) = 1.031, $p$ > .05, $d$ = .095). Stimuli with male-pitched voices and masculine prosody ($M_{XMm}$ = 3.570, $SD_{XMm}$ = .629) were rated significantly more competent than stimuli with male-pitched voices and feminine prosody ($M_{XMf}$ = 3.397, $SD_{XMf}$ = .635) ($t$(96) = 3.604, $p$ < .001, $d$ = .274). Once again, participants seemed to have a preference for voices with congruent gender traits. For a graph of this interaction, see *Figure 12.*

Finally, I ran post-hoc paired sample t-test for the interaction between face, pitch, and prosody. I found that face had an impact on competence ratings for all face-voice combinations except female-pitched voices with feminine prosody, which was only marginally significant. Stimuli with female-pitched voices and feminine prosody were rated marginally more competent when paired with male faces than when paired with female faces ($t(96) = 1.910$, $p < .1$, $d = .193$). Stimuli with female-pitched voices and masculine prosody were rated more competent when paired with female faces than male faces ($t(96) = 4.221$, $p < .001$, $d = .393$). Stimuli with androgynous-pitched voices and both masculine and feminine prosody were rated more competent when paired with male faces than when paired with female faces ($t_m(96) = 3.310$, $p < .01$, $d_m = .321$) ($t_f(\text{df}) = 3.793$, $p < .001$, $d_f = .425$). Stimuli with male-pitched voices and feminine prosody were rated significantly more competent when paired with female faces than male faces ($t(96) = 3.883$, $p < .001$, $d = .404$), while stimuli with male-pitched voices and masculine prosody were rated significantly more competent when paired with male-faces than female faces ($t(96) = 3.963$, $p < .001$, $d = .435$). In sum, except for male-pitched voices with feminine prosody and female pitched voices with masculine prosody, male faces were consistently rated more competent than female faces. For stimuli with incongruent pitch and prosody, male faces were rated significantly less competent than female faces. For graphs of this interaction, see *Figure 13*. For descriptive statistics, see *Table 10*.

In order to determine whether female, androgynous, and male categorized stimuli pairs differed on competence ratings, I ran another repeated measures ANOVA. There was a significant main effect of Gender category ($F(2) = 6.304$, $p < .01$, $\eta^2 = .062$). This was a moderate effect, as it accounted for 6.2% of the variance in competence ratings after excluding the proportion of variance due to other effects. Post-hoc paired sample t-tests revealed that female categorized stimuli ($M_F = 3.291$, $SD_F = .616$) were rated significantly less competent than male ($M_M = 3.440$, $SD_M = .621$) ($t(96) = 3.170$, $p < .01$, $d = .241$) and androgynous ($M_A = 3.464$,

$SD_A$ = .677) categorized stimuli ($t(96) = 2.983$, $p < .05$, $d = .267$). Male categorized and androgynous categorized stimuli were rated equally competent ($t(96) = .453$, $p > .05$, $d = .037$).

## Discussion

### Continuous and Binary Gender

Based on previous research, I hypothesized that the the pitch of a person's voice would be the most salient gendered cue for binary gender categorization. However, results from this study found that stimuli were categorized based on the gender of the face, not pitch, for all conditions except when the pitch and prosody were congruent with each other, but incongruent with the face ($F_{Mm}$, $M_{Ff}$). It is possible that this discrepancy from the previous research occured because the audio clips did not play automatically, thus respondents were exposed to the face before receiving any other gendered cues. The primacy effect of the faces' gender likely had an impact on the perceived gender of the stimuli (Anderson & Barrios, 1961); if all gendered cues were presented simultaneously I would expect my results to be more similar to previous research on gender categorization. Because the face and voice for $F_{Mm}$ and $M_{Ff}$ were so incongruent, this may have been powerful enough to overcome the primacy effect of face for these stimuli categories.

While face appeared to play the most significant role in binary gender categorization, it had no significant role in continuous gender ratings. As I hypothesized, pitch and prosody both played a significant role in continuous gender ratings, while facial cues did not. I predicted that pitch, rather than prosody, would be the most salient cue for continuous gender ratings. Instead, I found that pitch and prosody had equally large effects on continuous gender ratings; masculine prosody and male-pitched voices resulted in higher masculinity ratings than feminine prosody and androgynous- and female-pitched voices. I also expected an individual's endorsement of the gender binary to moderate these ratings, but ultimately found no significant effect.

One interesting finding was that stimuli that were categorized as feminine men were rated less masculine than stimuli that were categorized as masculine women. While it is conceivable that people perceive masculine women to be more masculine than feminine men, it is also possible that binary gender categorizations effected continuous gender ratings. For example, instead of rating a masculine woman's masculinity relative to all other people, participants might have rated the stimuli relative to all other women. If male and female categorized stimuli were rated on distinct scales, then that explains why face did not have a main effect on continuous gender. In other words, if all female categorized stimuli were rated from 1 to 5 relative to a typical female and all male categorized stimuli were rated from 1 to 5 relative to a typical male, then we likely would not see a meaningful difference between ratings for male and female faces, even if one existed. This potential discrepancy warrants future research on our perceptions of masculinity and femininity.

**Competence**

For competence, I hypothesized that stimuli with male-pitched voices and masculine prosody would be rated more competent than stimuli with male-pitched voices and feminine prosody, regardless of the gender of the face. Additionally, I predicted that female-pitched and androgynous-pitched voices with masculine prosody would be rated more competent than female- and androgynous-pitched voices with feminine prosody (especially when paired with male faces). The main effects of face and pitch confirm that the more "male" a face or voice is, the more competent it is perceived; however, the main effect of prosody suggested the opposite (feminine voices were rated more competent than masculine voices). It is likely that the main effect of prosody is driven primarily by the very low competence ratings for female-pitched masculine prosody voices. In cases where the only difference between the stimuli being compared was the gender of the face *or* the gender of the pitch, stimuli with a male face *or* male-pitch were rated more competent.

Stereotype incongruent male face stimuli were considered less competent than both stereotype congruent male face stimuli and stereotype incongruent female face stimuli, but more competent than stereotypical female face stimuli. This suggests that people who are perceived as stereotype incongruent men are considered less competent, but only relative to masculine males and stereotype incongruent stimuli that is perceived as female. Similarly, while stereotype incongruent stimuli that are perceived as female are considered more competent than feminine female stimuli, it is still considered less competent than masculine male stimuli.

These results suggest that stereotype incongruence results in worse backlash for male categorized stimuli than female categorized stimuli. It also suggests that "non-passing" transfeminine people (i.e. transfeminine people who are perceived as male) are considered less competent than masculine men. "Passing" transfeminine people (i.e. transfeminine people who are perceived as female) are also likely be considered more competent than non-passing transfeminine people, but are still considered less competent than masculine men.

However, when assessing the effects of the stereotype congruent/incongruent stimulus competence ratings, it is also important to recognize that there was very little variability in competence ratings in responses to this survey as a whole. All of the mean competence ratings were between 2.80 and 3.75. While there were still statistically significant differences in these ratings, the lack of variability among the stimuli limits my ability to draw confident conclusions. There are several potential reasons why the variability was so limited. First, it is possible that the bias against choosing the "extremes" on Likert scales (as mentioned above regarding *Task 2* of the pilot study, Hui & Triandis, 1989) in and of itself limited the variability of the scale (the majority of the ratings were between 2 and 4). Another potential confounding factor might have been the content of the sentences. While the sentences were controlled for affect, they were not controlled for the kind of content they contained. Some sentences were narratives, while others stated facts or formal definitions. It's reasonable to expect facts and definitions to receive higher competence ratings than narrative sentences across all other conditions, simply because the content is more intellectual. If this survey is to be replicated in the future, investigators should control for the "competence" of the sentence content as well.

**Warmth**

I hypothesized that warmth ratings would be lower for stereotype incongruent stimuli than congruent stimuli, and that stimuli with androgynous-pitched voices would be rated more warm than stereotype incongruent stimuli, but less warm than stereotype congruent stimuli. The face by pitch and pitch by prosody interactions confirm that stereotype incongruent stimuli received lower warmth ratings. For the face by pitch interaction, male faces with male-pitched voices were rated marginally more warm than female faces with male-pitched voices, while female faces with female-pitched voices were rated more warm than male faces with female-pitched voices. For the pitch by prosody interaction, female-pitched voices with feminine prosody were rated more warm than female pitched voices with masculine prosody, while male-pitched voices with masculine prosody were rated more warm than male-pitched voices with feminine prosody. On the other hand, the face by prosody interaction yielded results that were the opposite of what I predicted; female faces with feminine prosody were rated *less* warm than male faces with feminine prosody. It is possible that the main effects of face and prosody might explain this, as male faces were overall rated more warm than female faces and feminine prosody was overall rated more warm than masculine prosody. This would also explain why, in the face by pitch by prosody interaction, stimuli with androgynous pitched voices were rated more warm when paired with a male face than a female face, regardless of prosody.

Furthermore, while the findings about the main effect of pitch (female-pitch was rated more warm than both male and androgynous pitch, male-pitch was rated more warm than androgynous-pitch) confirm the hypothesis that androgynous-pitched voices would be rated less warm than more easily categorized voices (male- and female-pitched), nothing in the other interactions or effects conclusively suggested that ambiguous stimuli was consistently rated more or less warm than either stereotype congruent or incongruent stimuli.

The most interesting finding about warmth ratings, however, was one that I did not predict. When pitch is the only incongruent cue ($M_{Fm}$, $F_{Mf}$), stimuli are consistently rated less warm than when there are no incongruent cues or when face or prosody is the incongruent cue. In other words, people like pitch to have a match. It is unclear whether the preferred match was face or prosody, but in all cases, when pitch stood apart from both, the stimuli received lower warmth ratings.

This finding has important implications for transfeminine people specifically. While hormone replacement therapy for transfeminine people effects many physical aspects of their bodies, allowing them to be perceived as stereotypical women, the deepening of a person's voice during testosterone driven puberty cannot be reversed with estrogen. Therefore, the female face, male pitch, feminine prosody category maps very well onto the audio visual cues that many transfeminine people provide when they interact with the world. The reason this is not the same for transmasculine people who take testosterone is because testosterone causes both a change in face and in voice, making it highly uncommon to encounter an individual with a male face and distinctly female pitched voice. In addition to the fact that femininity is implicitly considered “subordinate” to masculinity in most of human society (see *hegemonic masculinity*, Connell, 1995), the permanency of voice change is one of the primary reasons why transfeminine people have worse outcomes than transmasculine people (James et al., 2016). In other words, transfeminine people are simply more visible than transmasculine people. Stimuli with a pitch that was incongruent with both face and prosody received the most severe backlash in warmth ratings, paralleling the prevalent negative backlash that transfeminine people experience everyday.

Overall, stereotype incongruent stimuli were also *usually* rated less warm than congruent stimuli, suggesting that visibly transgender and gender nonconforming people experience more backlash than cisgender and gender conforming people.

**Conclusions**

Overall, stereotype incongruent stimuli were seen as less warm than stereotype congruent stimuli, regardless of the gender they were perceived to be. However, this is not the case for competence ratings. Whether or not stereotype incongruent stimuli were considered more or less competent than stereotype congruent stimuli depended on whether those stimuli were perceived as male or female. To tie this back to the stereotype content model, these results suggest that female-passing transfeminine people are considered relatively low on warmth, but not competence, while non-passing transfeminine people are considered low on both warmth and competence. I correctly predicted that the stimuli that paralleled non-passing transfeminine people would be rated low on both warmth and competence, but I did not expect that the stimuli that paralleled female-passing transgender people would not reflect negative backlash in competence ratings. These findings put transfeminine people somewhere in the contemptuous to envious stereotype areas of the stereotype content model. While it is harder to place transmasculine people on the stereotype content model because they are frequently indistinguishable from cisgender men, the results suggest that transmasculine people who do not "pass" as male (i.e. female face, androgynous/female voice, masculine prosody) are rated relatively low on warmth and competence. I also did not anticipate that this stimulus *would* experience negative backlash in competence ratings. While I accurately hypothesized that the stimuli that paralleled transgender experiences would be subject to envious and contemptuous prejudice, I incorrectly predicted which stimuli would fall in each of these categories. I had expected the stimuli representing non-passing transmasculine people to be subject to envious prejudice and the stimuli representing transfeminine people to be subject to contemptuous prejudice. Instead, I found that transfeminine stimuli were subject to both kinds of prejudice, depending on whether or not the stimuli was perceived as female, and non-passing transmasculine stimuli were subject to contemptuous prejudice. For a graph of all of the stimuli categories plotted on the stereotype content model, see *Figure 14*.

Given the kinds of discrimination and prejudice that transgender people face, these designations within the stereotype content model make sense. Transgender people, and especially transfeminine people, are often considered undeserving of any kind of substantive government assistance or charity. Instead of recognizing that the majority of the negative outcomes that transgender people face are a direct result of this negative backlash and prejudice, many people have adopted the view that transgender people are “mentally ill and need serious treatment” (Payne, 2016). The idea that transgender people are not fit for society and do not deserve access to employment, transition (and non-transition) related health care, and many other important resources is highly characteristic of contemptuous prejudice.

Moreover, transgender people are also often denied rights and a visible place in society because they are scapegoated for problems they do not actually perpetuate. For example, “Bathroom bills,” such as North Carolina’s House Bill 2, scapegoat transgender people for gender-based violence, even though transgender people experience gender-based violence at higher rates than any other population (James et al., 2016). These bills make it illegal for transgender people to use restrooms that do not correspond with their assigned sex. Proponents of these bills often argue that it is “inappropriate for men to use the women’s bathroom” and that allowing transfeminine people to enter women’s spaces puts cisgender women “in harm’s way.” Not only does this invalidate the identities of transgender people, but it also shifts the onus of responsibility for gender-based violence from (primarily) cisgender men on to transgender people.

Another example of this kind of scapegoating is President Donald Trump’s claim and subsequent policy recommendation that transgender people are a “disruption” and a “burden” on the U.S. military (realDonaldTrump, 2017) and should therefore not be allowed to serve. Transgender people’s medical costs are a negligible portion of the military’s enormous budget, but because transgender people are an easy target, they are the ones who get blamed and punished for America's budget problems.

Transgender people challenge deeply embedded ideas about sex, gender, and gender roles. This can be threatening to the members of the dominant group (especially cisgender men) because it challenges their belief in the inherent superiority/inferiority of certain social groups (*social dominance orientation*, Sidanius & Pratto, 1999). In addition to the contemptuous prejudice that transgender people face, this threat to established social hierarchies ultimately drives the envious prejudice that transgender people experience. Again, while there is no data in this study to assess whether "passing" transmasculine people are subject to the same stereotypes, the similar negative outcomes that they experience suggest that, to an extent, they are.

Of course, it is important to recognize the limitations of this research. One limitation was the authenticity of the stimuli. Because the voices were paired with static images of faces, participants were told to imagine that the voice they heard belonged to the person whose face they were seeing. This was easier to do with some face-voice combinations than it was with others because some of the voices were so incongruent with the faces that it was difficult to imagine them coming from the same person. Similarly, in real life, we do not just interact with faces and voices. Faces and voices come with bodies, and together, these contribute to our perceptions of other people. While the mere association of the face and voice still had an effect, the results would likely be much more profound if the faces and voices were better integrated with each other using dynamic, embodied stimuli. Future studies could use computer generated imagery (CGI), advanced video editing technology, or actual people in order to create more realistic stimuli and gain a more accurate understanding of people's reactions to stereotype incongruent individuals in real life.

In addition to the presentation of the stimuli, many of the voice stimuli lost audio quality as the pitch was transposed. This corruption of the audio footage made some of the voices sound robotic and unrealistic, which probably had an impact on people's impressions of the stimuli. Thus, if this study is to be replicated, it will be important to ensure that the voice stimuli is convincing at all pitch levels.

Further, as addressed above, this study did not have stimuli that paralleled the transmasculine experience. It is generally more difficult to represent this in a controlled experimental study, as transmasculine people who have taken testosterone usually do not physically appear to be transgender. While male-passing transmasculine people are still subject to prejudice and discrimination, it is often only upon (voluntary or involuntary) disclosure of their identities. Thus, this research is not generalizable to male-passing transmasculine people and future research will need to find an alternative way to assess attitudes towards transmasculine people.

It is also important to point out that this study exclusively used white faces for the face stimuli in order to control for race. Given the already complex design of this research, adding race as a fourth independent variable was unrealistic. However, transgender people of color are often subject to much more severe outcomes than white transgender people (especially black and indigenous transgender people) (James et al., 2016), therefore, it's important for future studies to incorporate race and other intersectional identities into the research design.

Additionally, this study did not fully address the effect that ambiguity has on negative backlash, prejudice, and discrimination. While some stimuli were paired with androgynous-pitched voices, the distinctly male/female faces with which they were paired provided a salient gender cue that allowed for easy categorization. A future study could adopt a similar design as this study in order to assess whether easily categorized stimuli are considered more warm or competent than ambiguous stimuli. In order to do so, androgynous-pitched voices should be paired with androgynous faces, so that prosody is the only remaining cue.

Finally, this study did not record reaction time for any of the stimuli ratings. Reaction time is often used as a proxy for attitude accessibility, suggesting that the less time it takes to make a judgement, the easier the judgement was to make (Lavrakas, 2008). This information would be useful for helping assess people's implicit attitudes, in addition to the more explicit attitudes reflected in their judgements. Presumably, if stereotype congruent stimuli are quickly rated as highly warm/competent and stereotype incongruent/ambiguous stimuli are quickly rated as low on warmth/competence, this suggests that respondents implicitly favor stereotype congruent stimuli. Likewise, if stereotype congruent stimuli are quickly rated as highly warm/competent, while stereotype incongruent/ambiguous stimuli are slowly rated as warm/competent, that still suggests that the positive attitudes towards stereotype congruent stimuli are more accessible than the positive attitudes directed towards incongruent/ambiguous stimuli.

Research about transgender discrimination is relatively new in all disciplines. There is a lot of interdisciplinary work that needs to be done before we have the kind of understanding about transphobia that we do about sexism and racism. While this study attempted to set the foundation for research on transgender discrimination and prejudice, it hardly scratched the surface. This field of research is abounding with opportunities for further study.

## Acknowledgements

I would like to thank everyone who has helped me academically and emotionally throughout this research project. Thank you to Paul Thibodeau and Sara Verosky for serving on my honors committee and lending me their expertise in cognitive psychology, impression formation, and face research. Thank you to Max Kramer for helping to professionally record voice actors to use as stimulus in this study. Thank you to Joanie Gleason for keeping everything organized and helping me secure funding for this research. Thank you to all the other faculty in the Oberlin Department of Psychology who I consulted during this process. Finally, an enormous thank you to Cindy Frantz for making yourself available whenever I needed help, meeting with me week after week for over a year to put this project together, and for believing in me. I could not have done this without you and I am so glad to have had you as my mentor.

References

Adorno, T.W., Frenkel-Brunswik, E., Levinson, D.J., Sanford, R.N. (1950). *The Authoritarian Personality*. New York: Harper & Row.

Anderson, N. H., & Barrios, A. A. (1961). Primacy Effects in Personality Impression Formation. *Journal of Abnormal and Social Psychology*, 63(2), 346-350.

Articulation. (n.d.). In *English Oxford Dictionary online*. Retrieved from https://en.oxforddictionaries.com/definition/articulation

Bless, H., Schwarz, N., Bodenhausen, G. V., & Thiel, L. (2001). Personalized versus Generalized Benefits of Stereotype Disconfirmation: Trade-offs in the Evaluation of Atypical Exemplars and Their Social Groups. *Journal of Experimental Social Psychology*, 37, 386-397.

Block, C. *Feminine Voice*. Retrieved from http://www.nyspeechandvoicelab.net/transgender/voice-feminization/

Bussey, K., & Bandura, A. (1999). Social Cognitive Theory of Gender Development and Differentiation. *Psychological Review*, 106, 676-713.

Connell, R. W. (1995). *Masculinities*. Cambridge, U.K.: Polity Press.

Diamond, M. (2004). Sex, Gender, and Identity over the Years: A changing perspective. *Child and Adolescent Psychiatric Clinics of North America*, 13, 591-607.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A Model of (Often Mixed) Stereotype Content: Competence and Warmth Respectively follow from Perceived Status and Competition. Journal of Personality and Social Psychology, 82(6), 878-902.

Fiske, S. T. (2013). *Social beings: Core motives in social psychology* (3rd ed.). John Wiley & Sons.

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology*, Vols. 2 (4th ed.) (pp. 357-411). New York: McGraw-Hill.

Förster, J., Higgins, E. T., & Werth, L. (2004). How threat from stereotype disconfirmation triggers self-defense. *Social Cognition*, 22 (1), 54-74.

Freeman, J.B., & Ambady, N. (2011). When two become one: Temporally dynamic integration of face and voice. *Journal of Experimental Social Psychology*, 47 (1), 259-263.

Gelfer, M. P., & Bennett, Q.E. (2013). Speaking Fundamental Frequency and Vowel Formant Frequencies: Effects on Perception of Gender. *Journal of Voice*, 27 (5), 556-566.

Hui, C. H., & Triandis, H. C. (1989). Effects of Culture and Response Format on Extreme Response Style. *Journal of Cross-Cultural Psychology*, 20 (3), 296-309.

Human Rights Campaign Staff (2015, February 1). HRC National Survey of Likely Voters. Retrieved April 8, 2017, from http://www.hrc.org/resources/hrc-national-survey-of-likely-voters.

Intonation. (n.d.). In *Encyclopaedia Britannica online*. Retrieved from https://www.britannica.com/topic/intonation

James, S. E., Herman, J. L., Rankin, S., Keisling, M., Mottet, L., & Anafi, M. (2016). *The Report of the 2015 U.S. Transgender Survey*. Washington, DC: National Center for Transgender Equality.

Lavarkas, P. J. (2008). Response Latency. In *Encyclopedia of Survey Research Methods*.

Retrieved from http://methods.sagepub.com/reference/encyclopedia-of-survey-research-methods/n487.xml

Marques, J.M., Paez, D. & Abrams, D. (1998). Social identity and intragroup differentiation as subjective social control. In S. Worchel, J.F. Morales, D. Paez & J. -C. Deschamps (Eds.) *Social Identity: International perspectives* (124-141). Nova Iorque: Sage.

Martin, C. L. (1990). Attitudes and Expectations about Children with Nontraditional and Traditional Gender Roles. *Sex Roles*, 22 (3/4), 151-165.

Payne, D. (2016, July 7). The Transgender Suicide Rate Isn't Due To Discrimination. *The Federalist*. Retrieved from http://thefederalist.com/2016/07/07/evidence-the-transgender-suicide-rate-isnt-due-to-discrimination/

Plaks, J. E., Grant, H., Dweck, C.S. (2005). Violations of Implicit Theories and the Sense of Prediction and Control: Implications for Motivated Person Perception. *Journal of Personality and Social Psychology*, 88 (2), 245-262.

Prosody. (21 Feb 2018). In *Merriam-Webster online*. Retrieved from https://www.merriam-webster.com/dictionary/prosody

Rudman, L. A., & Fairchild, K. (2004). Reactions to counterstereotypic behavior: the role of backlash in cultural stereotype maintenance. *Journal of Personality and Social Psychology*, 87, 157–176.

Rudman, L. A., & Phelan, J. E. (2008). Backlash effects for disconfirming gender stereotypes in organizations. *Research in Organizational Behavior*, 28, 61–79.

Sidanius, J. & Pratto, F. (1999). Social Dominance. Cambridge, UK: Cambridge University Press.

realDonaldTrump. (2017, July 26). ....victory and cannot be burdened with the tremendous medical costs and disruption that transgender in the military would entail. Thank you." [Twitter Post]. Retrieved from https://twitter.com/realDonaldTrump/status/890197095151546369

Watson, R., Charest, I., Rouger, J., Casper, C., Latinus, M., & Belin, P. (2012). Audiovisual Integration of Face-Voice Gender Studied Using "Morphed Videos." In P. Belin, S. Campanella, & T. Ethofer (Eds.), *Integrating Face and Voice in Person Perception* (135-148). New York: Springer Science+Business Media.

Wood, W., & Eagly, A.H. (2002). A Cross-Cultural Analysis of the Behavior of Women and Men: Implications for the Origins of Sex Differences. *Psychological Bulletin*, 128 (5), 699-727.

## Tables

*Table 1*

| Voice | Target Gender | % Male | % Female | Mean CG (SD) |
|---|---|---|---|---|
| V1* | M | 86.2 | 13.8 | 4.59 (1.07) |
| V2* | F | 15.5 | 84.5 | 3.57 (1.38) |
| V3* | F | 6.9 | 93.1 | 2.53 (0.99) |
| V4 | M | 37.9 | 62.1 | 3.97 (1.08) |
| V5* | M | 31.0 | 69.0 | 3.79 (1.07) |
| V6* | F | 8.6 | 91.4 | 3.01 (1.33) |
| V7* | F | 1.7 | 98.3 | 2.21 (0.80) |
| V8 | M | 56.9 | 43.1 | 4.52 (0.99) |
| V9 | M | 44.8 | 55.2 | 4.71 (1.35) |

This table shows the target gender of each voice at the androgynous pitch frequency (165 Hz), as well as the percentage of the time that the voice was categorized as male and female. The furthest right column shows the average continuous gender rating and standard deviation for each voice. Grey text represents voices that were removed from the stimuli pool due to pilot testing results. Note: * = significant at $\alpha = .05$.

*Table 2*

| Face Gender | Male-Masculine (Mm) | Male-Feminine (Mf) | Andro-Masculine (Am) | Andro-Feminine (Af) | Female-Masculine (Fm) | Female-Feminine (Ff) |
|---|---|---|---|---|---|---|
| Male (M) | $M_{Mm}$ | $M_{Mf}$ | $M_{Am}$ | $M_{Af}$ | $M_{Fm}$ | $M_{Ff}$ |
| Female (F) | $F_{Mm}$ | $F_{Mf}$ | $F_{Am}$ | $F_{Af}$ | $F_{Fm}$ | $F_{Ff}$ |

This table shows the twelve combinations of faces (rows) and voices (columns) used in this study.

*Table 3: Binary Gender Descriptives*

| Stimulus | Mean | Std. Deviation |
|---|---|---|
| $F_{Ff}$ | 1.0000 | .00000[a] |
| $F_{Af}$ | .9948 | .05077 |
| $F_{Mf}$ | .9433 | .18925 |
| $F_{Fm}$ | .8608 | .29530 |
| $F_{Am}$ | .7887 | .31321 |
| $F_{Mm}$ | .5464 | .48457 |
| $M_{Ff}$ | .4639 | .47460 |
| $M_{Af}$ | .3969 | .46736 |
| $M_{Mf}$ | .1649 | .33635 |
| $M_{Fm}$ | .2526 | .40904 |
| $M_{Am}$ | .0979 | .25657 |
| $M_{Mm}$ | .0052 | .05077 |

This table shows the mean binary gender scores and standard deviation for each stimulus combination (0 = *always male*, 1 = *always female*). Each mean represents the percent of the time each stimulus was coded male or female. The first letter in the stimulus category names represents the face of the stimulus (F, M), the second letter represents the pitch of the voice (F, A, M), and the third letter represents the prosody of the voice (F, M). Note: a = t cannot be computed because standard deviation is 0.

*Table 4: Binary Gender Test Statistics*

| | Test Value = .5 | | |
|---|---|---|---|
| Stimulus | t | df | p-value |
| $F_{Af}$ | 96.000 | 96 | .000 |
| $F_{Mf}$ | 23.070 | 96 | .000 |
| $F_{Fm}$ | 12.034 | 96 | .000 |
| $F_{Am}$ | 9.077 | 96 | .000 |
| $F_{Mm}$ | .943 | 96 | .348 |
| $M_{Ff}$ | -.749 | 96 | .456 |
| $M_{Af}$ | -2.173 | 96 | .032 |
| $M_{Mf}$ | -9.811 | 96 | .000 |
| $M_{Fm}$ | -5.957 | 96 | .000 |
| $M_{Am}$ | -15.434 | 96 | .000 |
| $M_{Mm}$ | -96.000 | 96 | .000 |

This table shows the t-statistics, degrees of freedom, and p-values for the one-sample t-tests comparing the mean binary gender score of each stimulus to an expected mean binary gender score of 0.5 (which represents the binary gender score if stimulus was categorized as male and female 50% of the time). The first letter in the stimulus category names represents the face of the stimulus (F, M), the second letter represents the pitch of the voice (F, A, M), and the third letter represents the prosody of the voice (F, M). Note: There is no t-statistic for $F_{Ff}$ because there was no variability in binary gender categorization for this stimulus.

*Table 5: Tests of Within-Subjects Effects for Continuous Gender Ratings*

| Source | df | Mean Square | F | Sig. | Partial Eta Squared |
|---|---|---|---|---|---|
| Face | 1 | .379 | .230 | .632 | .002 |
| Error(Face) | 96 | 1.645 | | | |
| Pitch | 2 | 205.193 | 407.774 | .000 | .809 |
| Error(Pitch) | 192 | .503 | | | |
| Prosody | 1 | 466.640 | 502.017 | .000 | .839 |
| Error(Prosody) | 96 | .930 | | | |
| Face * Pitch | 2 | 1.044 | 4.384 | .014 | .044 |
| Error(Face*Pitch) | 192 | .238 | | | |
| Face * Prosody | 1 | 2.235 | 9.171 | .003 | .087 |
| Error(Face*Prosody) | 96 | .244 | | | |
| Pitch * Prosody | 2 | 2.373 | 6.349 | .002 | .062 |
| Error(Pitch*Prosody) | 192 | .374 | | | |
| Face * Pitch * Prosody | 2 | 5.548 | 16.802 | .000 | .149 |
| Error(Face*Pitch*Prosody) | 192 | .330 | | | |

This table shows the degrees of freedom, mean squares, F-statistics, p-values, and effect sizes for the repeated measures ANOVA assessing the effects of face, pitch, and prosody on continuous gender ratings.

*Table 6: Continuous Gender Rating Descriptives*

| Stimulus | Mean | Std. Deviation |
|---|---|---|
| $F_{Ff}$ | 1.222 | .388 |
| $F_{Af}$ | 2.170 | .807 |
| $F_{Mf}$ | 2.742 | .907 |
| $F_{Fm}$ | 2.773 | .764 |
| $F_{Am}$ | 3.330 | .750 |
| $F_{Mm}$ | 4.093 | .836 |
| $M_{Ff}$ | 1.567 | .840 |
| $M_{Af}$ | 2.021 | .750 |
| $M_{Mf}$ | 2.701 | .735 |
| $M_{Fm}$ | 2.402 | .656 |
| $M_{Am}$ | 3.180 | .630 |
| $M_{Mm}$ | 4.242 | .582 |

This table shows the mean continuous gender ratings and standard deviation for each stimulus combination (1 = *very feminine*, 5 = *very masculine*). The first letter in the stimulus category names represents the face of the stimulus (F, M), the second letter represents the pitch of the voice (F, A, M), and the third letter represents the prosody of the voice (F, M).

*Table 7: Tests of Within-Subjects Effects for Warmth Ratings*

| Source | df | Mean Square | F | Sig. | Partial Eta Squared |
|---|---|---|---|---|---|
| Face | 1 | 78.354 | 193.594 | .000 | .669 |
| Error(Face) | 96 | .405 | | | |
| Pitch | 2 | 20.182 | 53.408 | .000 | .357 |
| Error(Pitch) | 192 | .378 | | | |
| Prosody | 1 | .003 | .007 | .933 | .000 |
| Error(Prosody) | 96 | .491 | | | |
| Face * Pitch | 2 | 25.395 | 79.439 | .000 | .453 |
| Error(Face*Pitch) | 192 | .320 | | | |
| Face * Prosody | 1 | 63.560 | 145.344 | .000 | .602 |
| Error(Face*Prosody) | 96 | .437 | | | |
| Pitch * Prosody | 2 | 7.902 | 21.927 | .000 | .186 |
| Error(Pitch*Prosody) | 192 | .360 | | | |
| Face * Pitch * Prosody | 2 | 69.932 | 230.626 | .000 | .706 |
| Error(Face*Pitch*Prosody) | 192 | .303 | | | |

This table shows the degrees of freedom, mean squares, F-statistics, p-values, and effect sizes for the repeated measures ANOVA assessing the effects of face, pitch, and prosody on warmth ratings.

*Table 8: Warmth Ratings Descriptives*

| Stimulus | Mean | Std. Deviation |
|---|---|---|
| $F_{Ff}$ | 3.655 | .690 |
| $F_{Af}$ | 2.897 | .732 |
| $F_{Mf}$ | 2.784 | .790 |
| $F_{Fm}$ | 2.768 | .711 |
| $F_{Am}$ | 2.448 | .694 |
| $F_{Mm}$ | 3.077 | .708 |
| $M_{Ff}$ | 3.758 | .778 |
| $M_{Af}$ | 3.119 | .695 |
| $M_{Mf}$ | 2.985 | .701 |
| $M_{Fm}$ | 2.423 | .647 |
| $M_{Am}$ | 2.912 | .711 |
| $M_{Mm}$ | 3.113 | .705 |

This table shows the mean warmth ratings and standard deviation for each stimulus combination (1 = *not at all warm*, 5 = *extremely warm*). The first letter in the stimulus category names represents the face of the stimulus (F, M), the second letter represents the pitch of the voice (F, A, M), and the third letter represents the prosody of the voice (F, M).

*Table 9: Tests of Within-Subjects Effects for Competence Ratings*

| Source | df | Mean Square | F | Sig. | Partial Eta Squared |
|---|---|---|---|---|---|
| Face | 1 | 1.701 | 5.008 | .028 | .050 |
| Error(Face) | 96 | .340 | | | |
| Pitch | 2 | 12.230 | 31.273 | .000 | .246 |
| Error(Pitch) | 192 | .391 | | | |
| Prosody | 1 | 4.150 | 7.266 | .008 | .070 |
| Error(Prosody) | 96 | .571 | | | |
| Face * Pitch | 2 | 4.016 | 11.281 | .000 | .105 |
| Error(Face*Pitch) | 192 | .356 | | | |
| Face * Prosody | 1 | .078 | .225 | .636 | .002 |
| Error(Face*Prosody) | 96 | .344 | | | |
| Pitch * Prosody | 2 | 10.027 | 32.139 | .000 | .251 |
| Error(Pitch*Prosody) | 192 | .312 | | | |
| Face * Pitch * Prosody | 2 | 7.437 | 29.799 | .000 | .237 |
| Error(Face*Pitch*Prosody) | 192 | .250 | | | |

This table shows the degrees of freedom, mean squares, F-statistics, p-values, and effect sizes for the repeated measures ANOVA assessing the effects of face, pitch, and prosody on competence ratings.

*Table 10: Competence Ratings Descriptives*

| Stimulus | Mean | Std. Deviation |
|---|---|---|
| $F_{Ff}$ | 3.371 | .733 |
| $F_{Af}$ | 3.418 | .728 |
| $F_{Mf}$ | 3.546 | .729 |
| $F_{Fm}$ | 3.144 | .860 |
| $F_{Am}$ | 3.376 | .890 |
| $F_{Mm}$ | 3.407 | .761 |
| $M_{Ff}$ | 3.521 | .816 |
| $M_{Af}$ | 3.747 | .817 |
| $M_{Mf}$ | 3.247 | .750 |
| $M_{Fm}$ | 2.820 | .788 |
| $M_{Am}$ | 3.655 | .846 |
| $M_{Mm}$ | 3.732 | .733 |

This table shows the mean warmth ratings and standard deviation for each stimulus combination (1 = *not at all competent*, 5 = *extremely competent*). The first letter in the stimulus category names represents the face of the stimulus (F, M), the second letter represents the pitch of the voice (F, A, M), and the third letter represents the prosody of the voice (F, M).

**Figures**

*Figure 1*

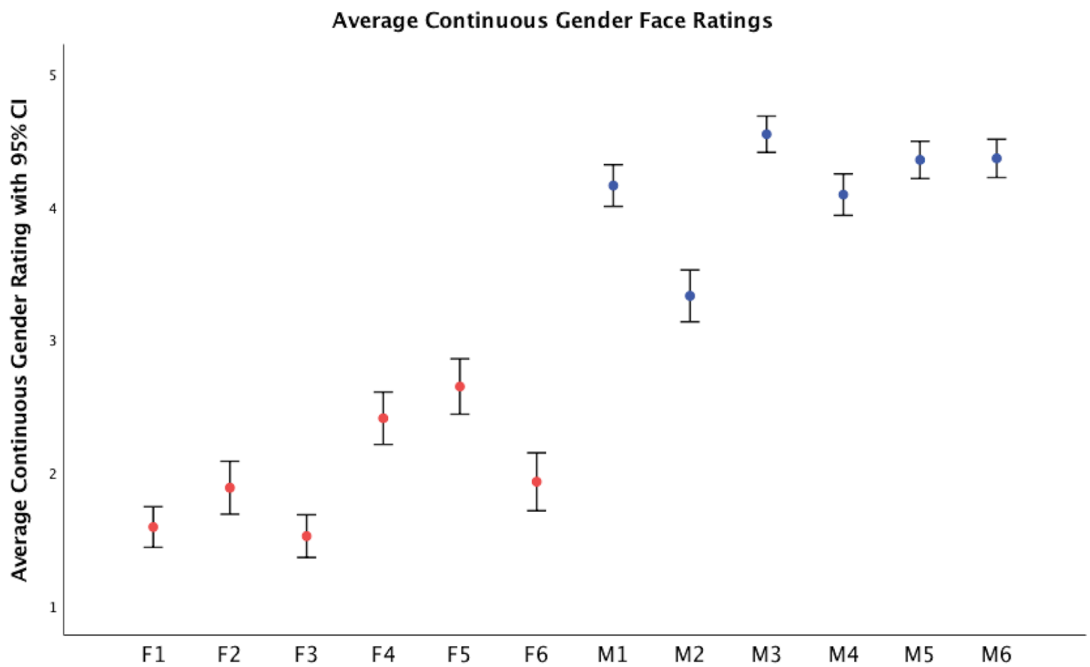



*Figure 1* shows the average continuous gender rating with a 95% confidence interval for each face in the pilot test. Red points represent female faces and blue points represent male faces. This graph shows that all female faces were rated significantly more less masculine than all male faces at $\alpha = .05$.

*Figure 2*

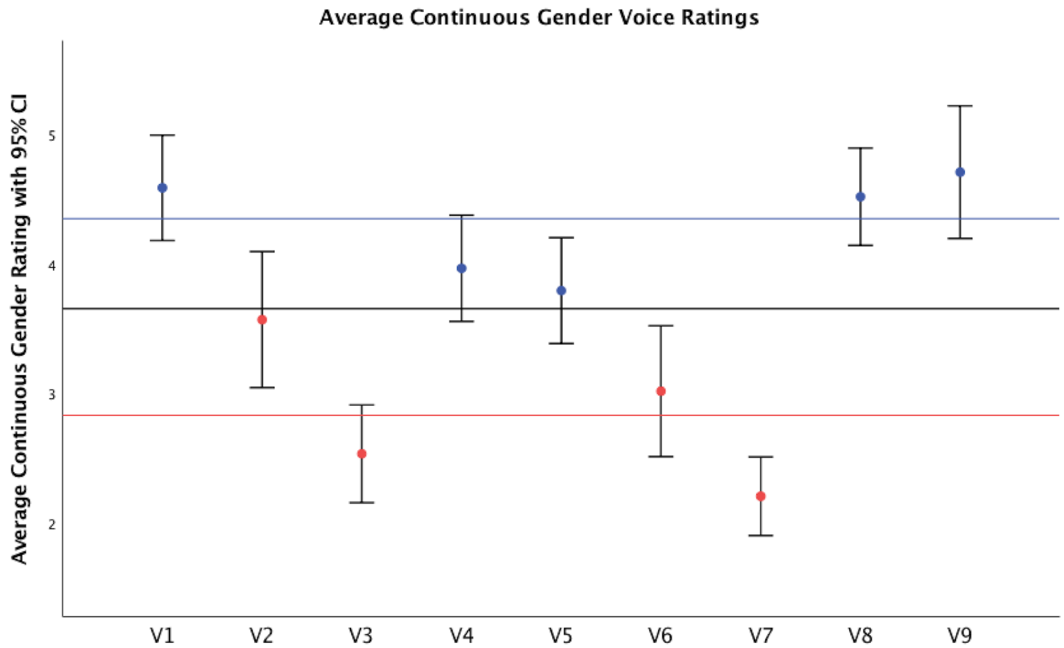



*Figure 2* shows the average continuous gender rating with a 95% confidence interval for each voice in the pilot test. Red points represent feminine voices and blue points represent masculine voices. The black line represents the overall mean continuous gender for all voices, while the red line represents the mean continuous gender for feminine voices, and the blue line represents the mean continuous gender for masculine voices. This graph shows that all feminine voices except V2 were rated significantly less masculine than the masculine voice mean and the overall mean ($\alpha = .05$). It also shows that all masculine voices except V4 and V5 were rated significantly more masculine than the feminine voice means and the overall mean ($\alpha = .05$). V2, V4, and V5 were removed from the stimulus pool for the main experiment to ensure that all vocal prosody was considered distinctly masculine or feminine.

*Figure 3*

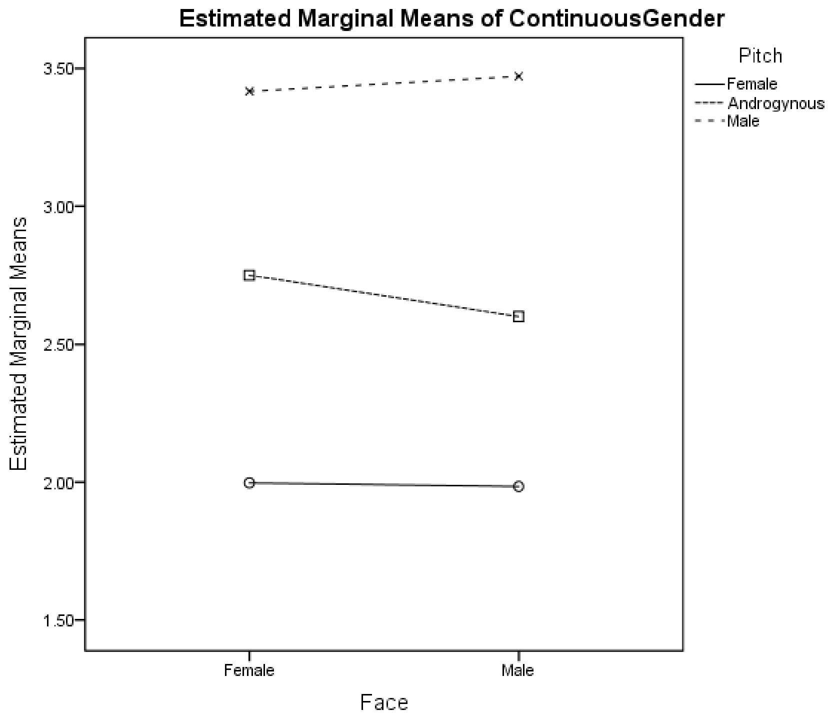



*Figure 3* displays the interaction between Face and Pitch for continuous gender ratings (0 = *very feminine*, 5 = *very masculine*).

*Figure 4*

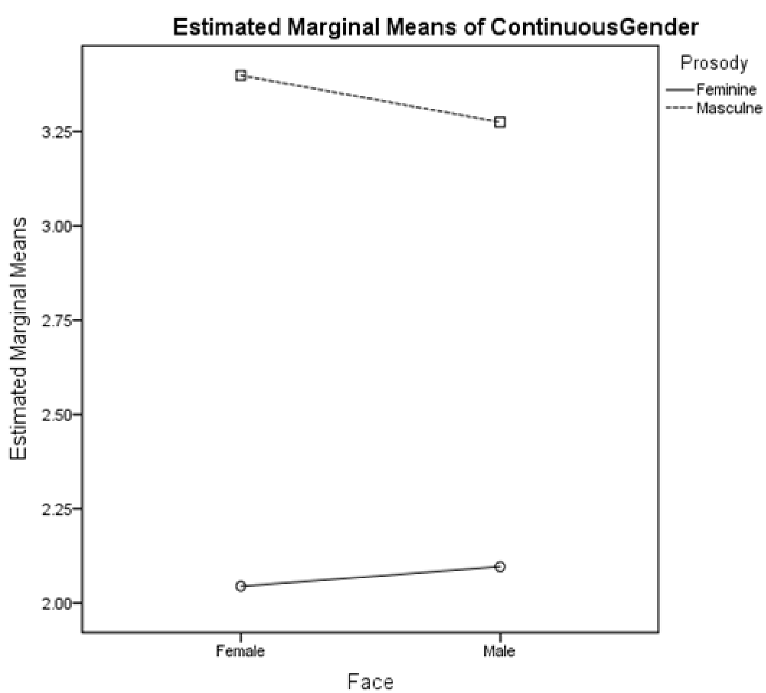



*Figure 4* displays the interaction between Face and Prosody for continuous gender ratings (0 = *very feminine*, 5 = *very masculine*).

*Figure 5*

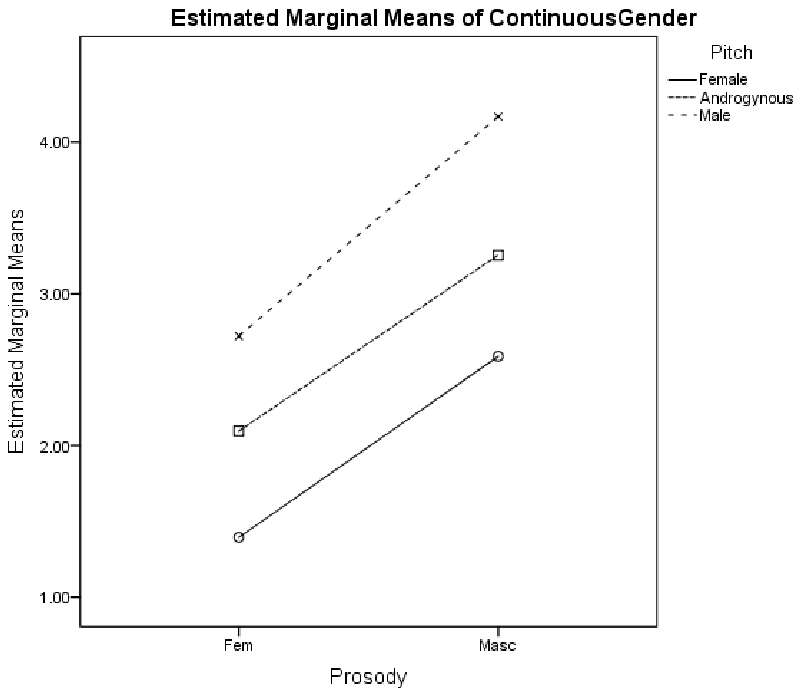



*Figure 5* displays the interaction between Prosody and Pitch for continuous gender ratings (0 = *very feminine*, 5 = *very masculine*).

*Figure 6A*

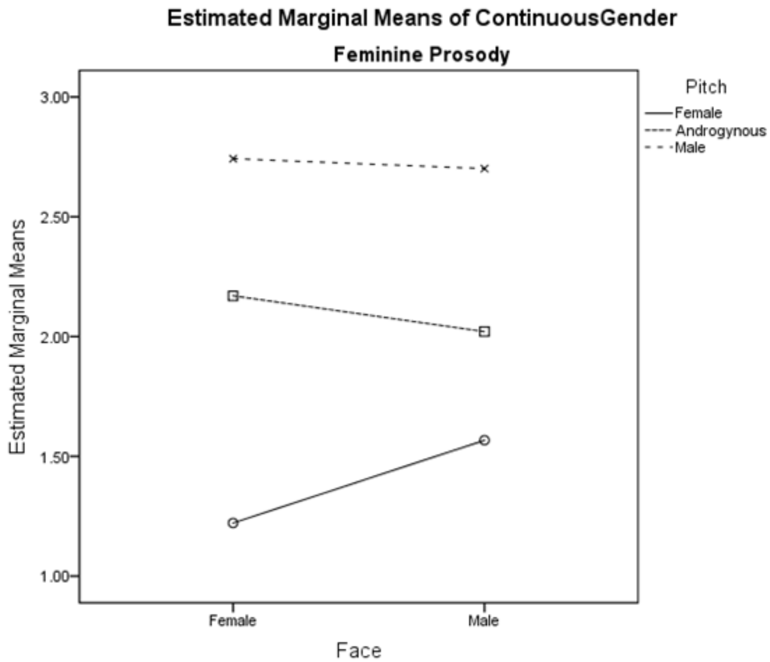



*Figure 6A* displays the interaction between Face and Pitch for feminine prosody voices for continuous gender ratings (0 = *very feminine*, 5 = *very masculine*).

*Figure 6B*

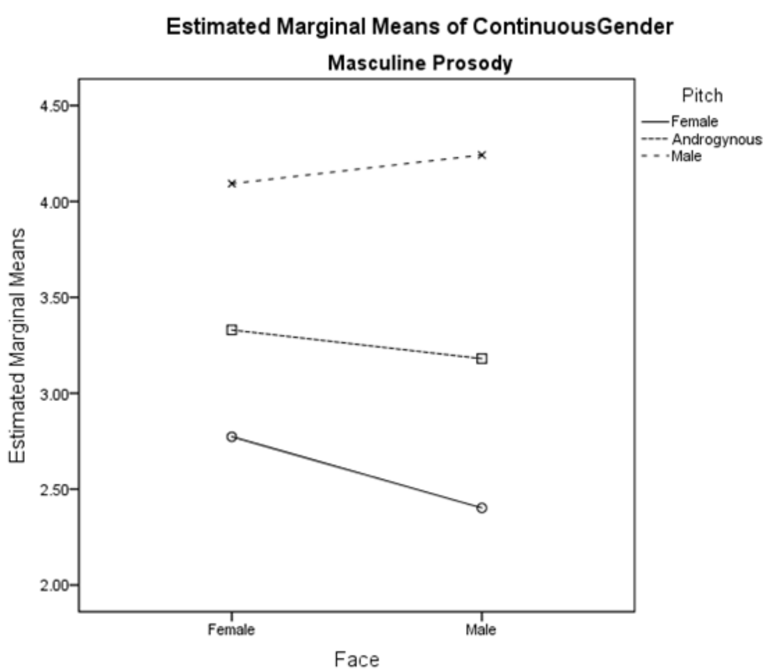



*Figure 6B* displays the interaction between Face and Pitch for masculine prosody voices for continuous gender ratings (0 = *very feminine*, 5 = *very masculine*).

*Figure 7*

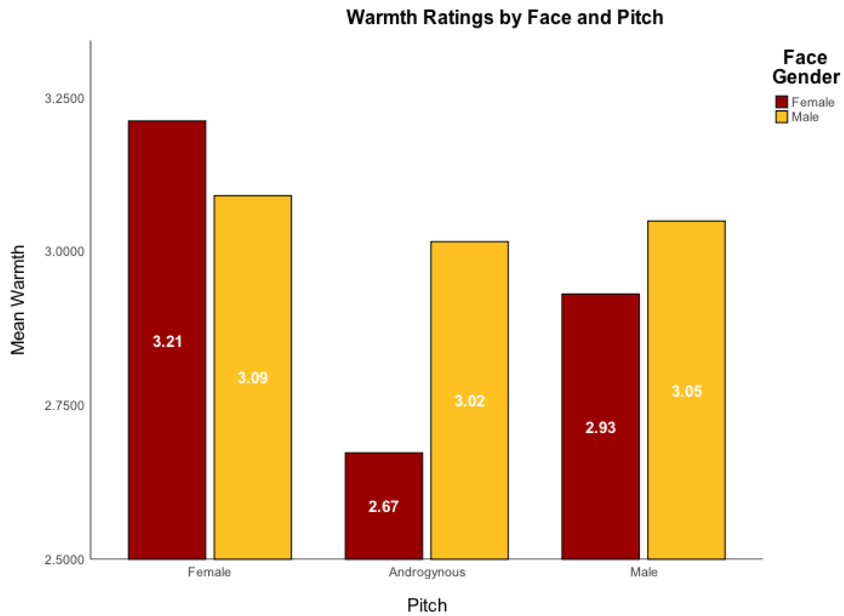



*Figure 7* shows the interaction between Face and Pitch for warmth ratings (1 = *not at all warm*, 5 = *extremely warm*).

*Figure 8*

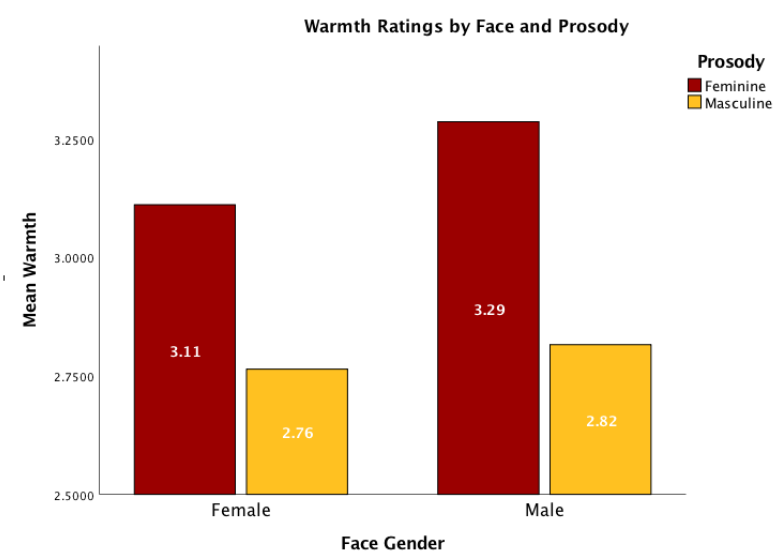



*Figure 8* shows the interaction between Face and Prosody for warmth ratings (1 = *not at all warm*, 5 = *extremely warm*).

*Figure 9*

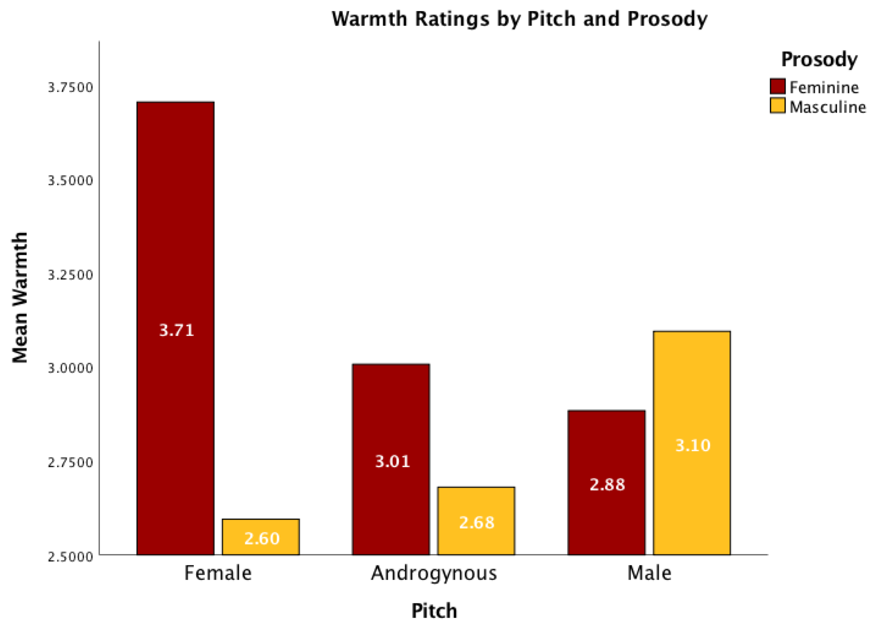



*Figure 9* shows the interaction between Prosody and Pitch for warmth ratings (1 = *not at all warm*, 5 = *extremely warm*).

*Figure 10A*

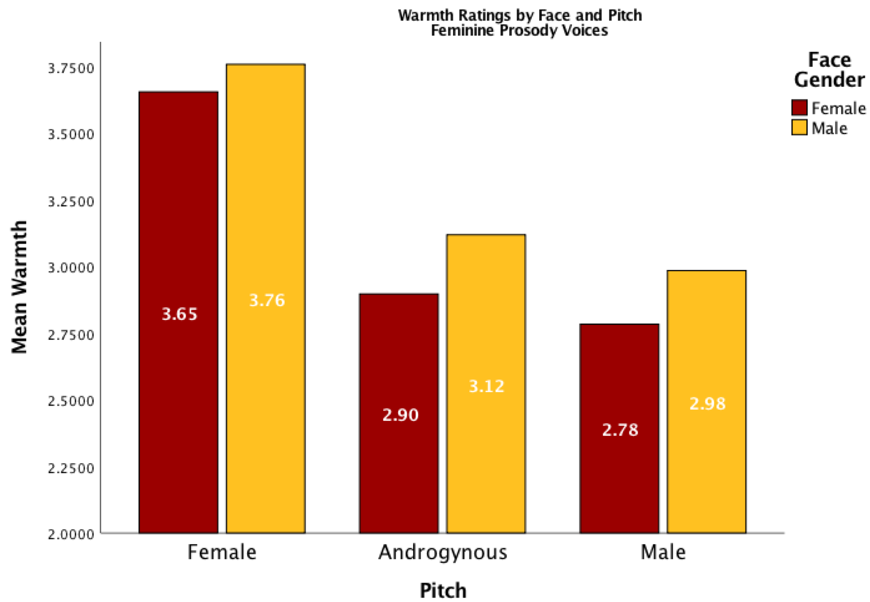



*Figure 10A* shows the interaction between Face and Pitch for feminine prosody voices for warmth ratings (1 = *not at all warm*, 5 = *extremely warm*).

*Figure 10B*

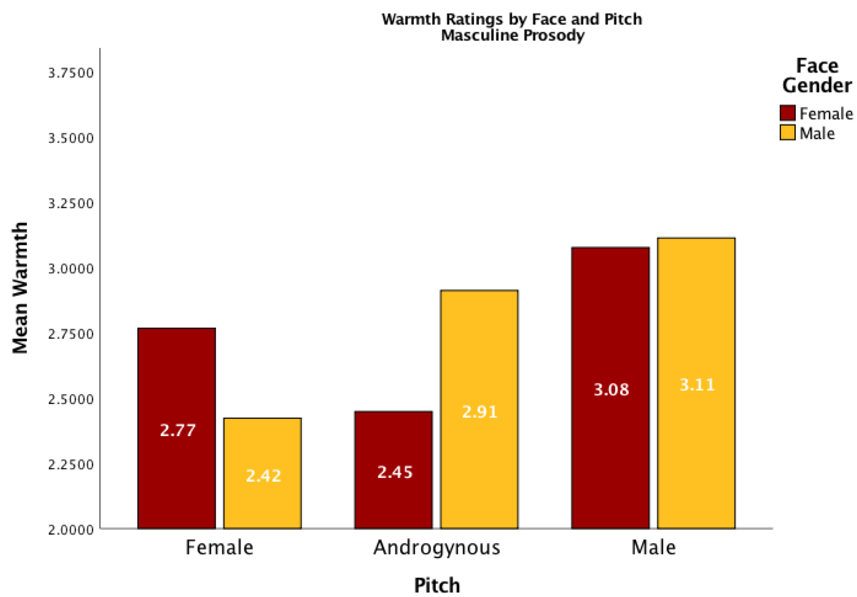



*Figure 10B* shows the interaction between Face and Pitch for masculine prosody voices for warmth ratings (1 = *not at all warm*, 5 = *extremely warm*).

*Figure 11*

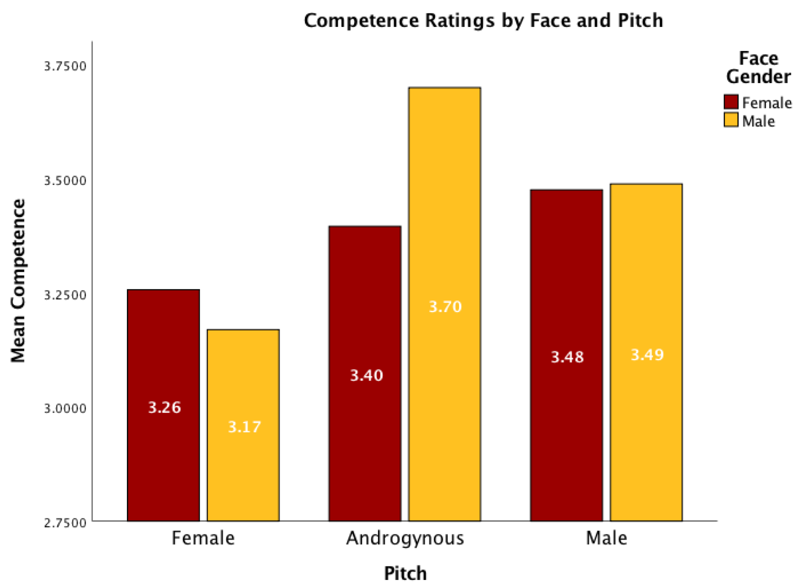



*Figure 11* shows the interaction between Face and Pitch for competence ratings (1 = *not at all competent*, 5 = *extremely competent*).

*Figure 12*

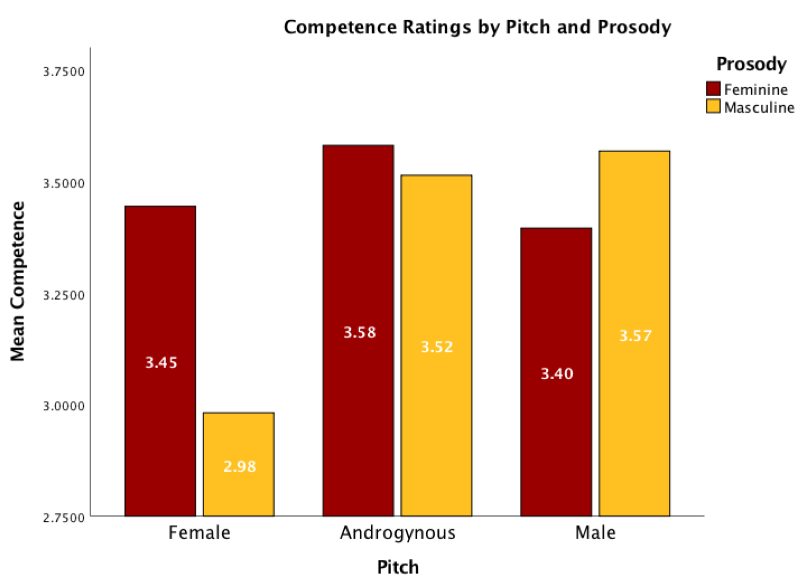



*Figure 11* shows the interaction between Prosody and Pitch for competence ratings (1 = *not at all competent*, 5 = *extremely competent*).

*Figure 13A*

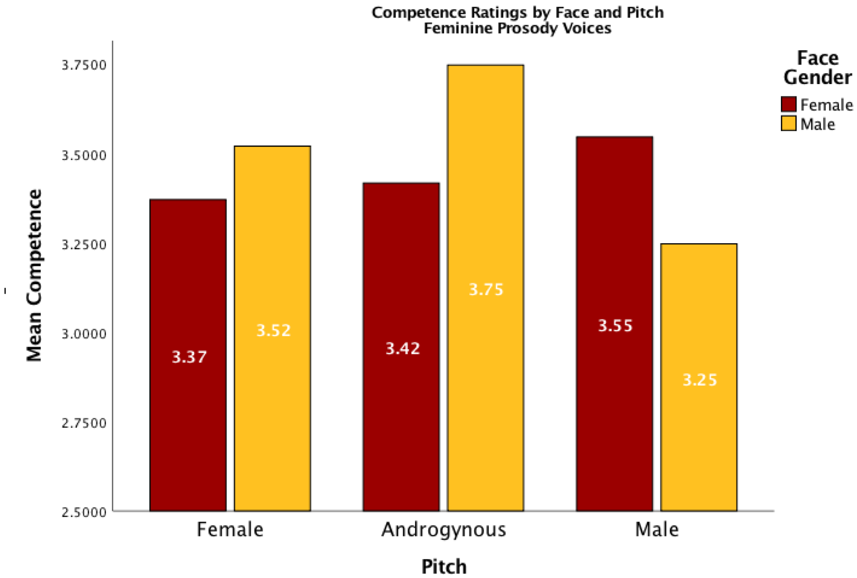



*Figure 13A* shows the interaction between Face and Pitch for feminine prosody voices for competence ratings (1 = *not at all competent*, 5 = *extremely competent*).

*Figure 13B*

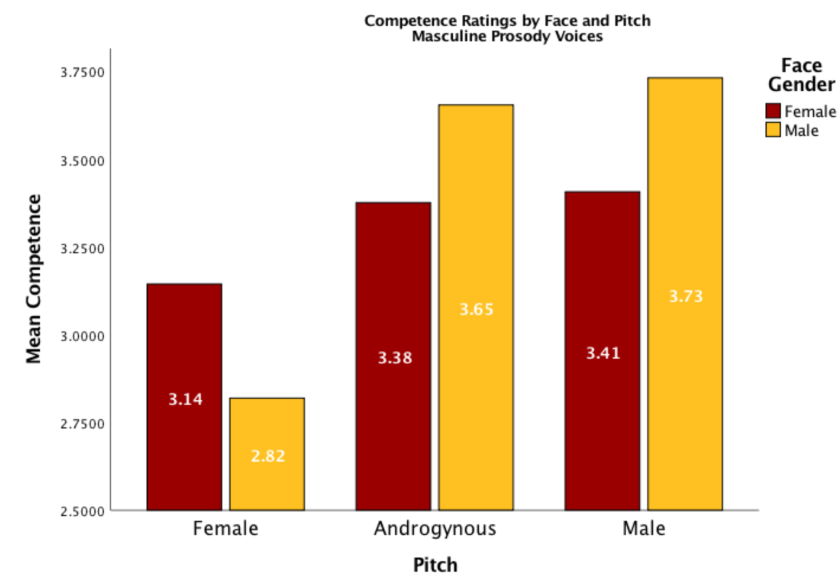



*Figure 13B* shows the interaction between Face and Pitch for masculine prosody voices for competence ratings (1 = *not at all competent*, 5 = *extremely competent*).

*Figure 14*

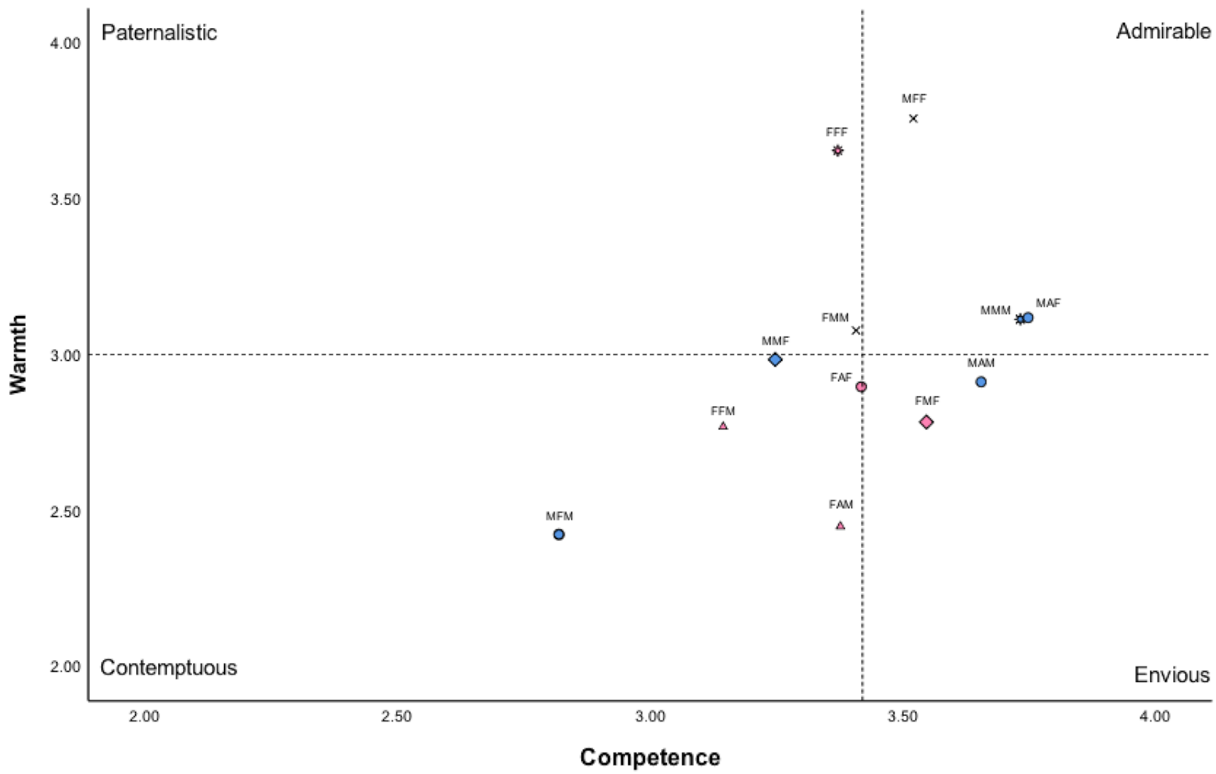
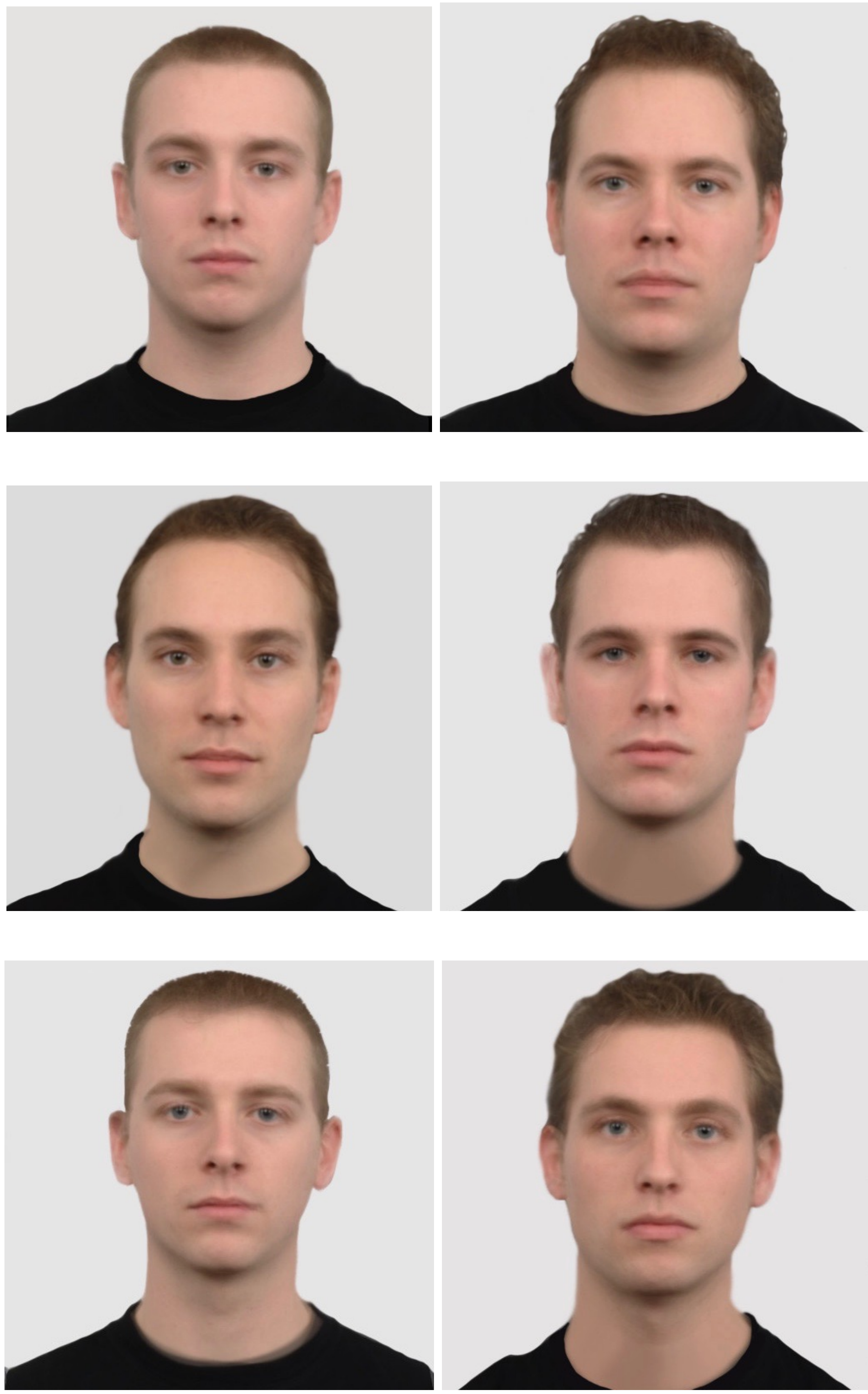



*Figure 14* shows each of the stimulus categories plotted based on their average warmth and competence ratings (lower ratings = less warm/competent), as well as the stereotypes and prejudice with which each quadrant is associated. The vertical line represents the mean competence rating ($x = 3.42$) while the horizontal line represents the mean warmth rating ($y = 2.99$). These lines make the four quadrants. Blue points were categorized as male and pink points were categorized as female. Stimuli combinations that were categorized as neither male or female are represented with an ‘x’. Stimuli categories that paralleled transfeminine gender presentation are represented using a diamond, stimuli that paralleled non-passing transmasculine gender presentation are represented with a triangle, and stimuli that paralleled cisgender-passing/gender conforming presentation are represented with a an 8-point star. Circles represent other stimuli that do not parallel any particular gender presentation.

# Appendix A

## Composite Faces



# Appendix B

## Neutral Sentences

1. William Shakespeare was an English poet, playwright, and actor, widely regarded as the greatest writer in the English Language and the world's most renowned dramatist.
2. Orogeny is a process in which a section of the earth's crust is folded and deformed to form a mountain range.
3. Mount Rainier is the highest mountain of the Cascade Range of the Pacific Northwest, and the highest mountain in the U.S. state of Washington.
4. Natural gas production remained relatively stable from the mid-1970s until the middle of the first decade of the twenty-first century.
5. All you really need to know is which test is appropriate for a particular situation and how to calculate its test statistic.
6. The Demand curve is a function that relates to the quantity of a commodity or service consumers wish to purchase based on the price of that commodity.
7. To prepare for hibernation, grizzly bears must prepare a den and consume an immense amount of food, as they do not eat during hibernation.
8. During my stay at Starkfield, I lodged with a middle-aged widow colloquially known as Mrs. Ned Hale.
9. One survey found that 53 percent of laboratories had stopped offering or developing at least one genetic test because of patent enforcement.
10. The Federal Policy for the Protection of Human subjects, also known as the Common Rule, requires informed consent for all human-subject research.
11. She stayed at home the next day, ate the leftover crab cakes Bill sent home with her, and talked to John on the phone.
12. The hens perched themselves on the window-sills, the pigeons fluttered up to the rafters, and the sheep and cows lay down behind the pigs and began to chew the cud.
13. Human migration has ranged from journeys of a few miles to epic travels across continents and oceans.
14. We crept through a broken hedge, groped our way up the path, and planted ourselves on a flower-pot under the drawing-room window.
15. The proterozoic is a geological eon representing the time just before the proliferation of complex life on earth.
16. A rift valley is a linear-shaped lowland between several highlands or mountain ranges created by the action of a geological rift or fault.
17. By mass, Iron is the most common element on earth, forming much of the earth's outer and inner core.
18. Pablo Picasso was a Spanish painter, sculptor, printmaker, ceramicist, stage designer, poet, and playwright who spent most of his adult life in France.
19. A Mars rover is an automated motor vehicle that propels itself across the surface of the planet Mars upon arrival.
20. In 1958, President Dwight D. Eisenhower established the National Aeronautics and Space Administration, which is more commonly referred to as NASA.
21. In 1969, Neil Armstrong, an American astronaut and engineer, became the first person to walk on the moon.
22. With a population of 21.3 million people, Mexico City is the largest metropolitan area in the western hemisphere.
23. Yellowstone National Park, the first National Park in the United States, is known for its wildlife and its many geothermal features.
24. The occipital lobe is the visual processing center of the mammalian brain, containing most of the anatomical region of the visual cortex.

# Appendix C

## Pilot Demographic Questionnaire

1. Which categories describe you? (Check all that apply)

    American Indian or Alaskan Native, Asian, Black or African-American, Hispanic or Latino, Middle Eastern or North African, Native Hawaiian or other Pacific Islander, White, Other (please specify)

2. How would you describe your sexual orientation?

    Heterosexual, Homosexual, Polysexual, Something Else

3. What is your gender?

    Male, Female, Androgynous, Something else

4. Do you identify as transgender?

    Yes, No

5. How would you describe your political orientation? Please respond using the scale provided.

    1 = Extremely Conservative, 7 = Extremely Liberal

6. How would you describe your household's financial situation? Please respond using the scale provided.

    1 = Economically Disadvantaged, 7 = Economically Well Off

## Appendix D

### Continuous Gender Post-Hoc Analyses

I ran post-hoc paired sample t-tests to assess the interaction between face and pitch. Initially, I ran tests comparing male and female faces within each pitch category. The tests revealed no significant differences between male and female faces for any of the pitch categories, but there were obvious discrepancies in the *change* in continuous gender ratings from male to female between the pitch categories. To examine the difference in the in change in continuous gender, I computed difference scores for each face-pitch combination. Paired sample t-tests revealed that change in continuous gender ratings for male-pitched voices ($M_{\Delta M}$ = -.054, $SD_{\Delta M}$ = .858) was significantly smaller than the change for androgynous-pitched voices ($M_{\Delta A}$ = .150, $SD_{\Delta A}$ = .956) ($t(96) = 2.869$, $p < .01$, $d$= .225). The change in continuous gender ratings for female-pitched voices ($M_{\Delta F}$ = .013, $SD_{\Delta F}$ = .687) was only marginally smaller for than androgynous pitched voices ($t(96) = 1.900$, $p < .1$, $d = .165$). The change in continuous gender rating for male-pitched vs. female-pitched voices was not statistically significant ($t(96) = .996$, $p > .05$, $d = .086$). For a graph of this interaction, see *Figure 3*.

Post-hoc paired sample t-tests for the face by prosody interaction showed that for both male and female faces stimuli with masculine prosody voices ($M_{MXm}$ = 3.275, $SD_{MXm}$ = .431, $M_{FXm}$ = 3.399, $SD_{FXm}$ = .603) were rated more masculine than feminine prosody voices ($M_{MXf}$ = 2.096, $SD_{MXf}$ = .569, $M_{FXf}$ = 2.045, $SD_{FXf}$ = .558). Like the face by pitch interaction, however, the change in continuous gender ratings between male and female faces was statistically significantly different. There was a greater change in continuous gender ratings between prosodies when they were paired with female faces ($M_{\Delta F}$ = -1.354, $SD_{\Delta F}$ = .667) than with male faces ($M_{\Delta M}$ = -1.179,

$SD_{\Delta M}$ = .580) ($t$(96) = 3.028, $p < .01$, $d$ = .280). In other words, prosody had a greater role in perceptions of masculinity/femininity when paired with female faces than when paired with male faces. For a graph of this interaction, see *Figure 4*.

Similar to the findings about the face by prosody interaction, the post-hoc paired sample t-tests for the pitch by prosody interaction revealed that masculine prosody voices ($M_{XFm}$ = 2.588, $SD_{XFm}$ = .525, $M_{XAm}$ = 3.255, $SD_{XAm}$ = .431, $M_{XMm}$ = 4.168, $SD_{XMm}$ = .573) were rated more masculine than feminine prosody voices ($M_{XFf}$ = 1.394, $SD_{XFf}$ = .500, $M_{XAf}$ = 2.095, $SD_{XAf}$ = .563, $M_{XMf}$ = 2.722, $SD_{XMf}$ = .594) across all levels of pitch. Again, I computed difference variables for the change in continuous gender rating based on pitch within all three pitch categories and then ran paired sample-tests to compare them. I found significant differences in the change in continuous gender between female- ($M_{\Delta F}$ = -1.193, $SD_{\Delta F}$ = .667) and male-pitched voices ($M_{\Delta M}$ = -1.446, $SD_{\Delta M}$ = .838) ($t$(96) = 2.525, $p < .05$, $d$ = .334) and androgynous- ($M_{\Delta A}$ = -1.160, $SD_{\Delta A}$ = .728) and male-pitched voices ($t$(96) = 3.649, $p < .001$, $d$ = .364), but not between female- and androgynous-pitched voices ($t$(96) = .402, $p > .05$, $d$ = .047). For a graph of this interaction, see *Figure 5*.

# Appendix E

## Post-screen questionnaire, reliability analyses, and scale correlations

### Transphobia Scale

| | Completely Disagree (1) | Disagree (2) | Somewhat Disagree (3) | Neither agree nor disagree (4) | Somewhat Agree (5) | Agree (6) | Completely Agree (7) |
|---|---|---|---|---|---|---|---|
| I don't like it when someone is flirting with me, and I can't tell if they are a man or a woman. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| I think there is something wrong with a person who says that they are neither a man nor a woman. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| I would be upset, if someone I'd known a long time revealed to me that they used to be another gender. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| I avoid people on the street whose gender is unclear to me. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| When I meet someone, it is important for me to be able to identify them as a man or a woman. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| I believe that the male/ female dichotomy is natural. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| I am uncomfortable around people who don't conform to traditional gender | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| roles, e.g., aggressive women or emotional men. | | | | | | | |
| I believe that a person can never change their gender. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| A person's genitalia define what gender they are, e.g., a penis defines a person as being a man, a vagina defines a person as being a woman. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

*Cronbach's alpha* = .917

**Endorsement of the Gender binary**

| | Completely Disagree (1) | Somewhat Disagree (2) | Neither agree nor disagree (3) | Somewhat Agree (4) | Completely Agree (5) |
|---|---|---|---|---|---|
| These days there is not enough respect for the natural divisions between the sexes. | ☐ | ☐ | ☐ | ☐ | ☐ |

**Authoritarianism**
(0 = *non-authoritarian*, 1= *authoritarian*)

Which is the most desirable quality for a child to have:

1. 
    a. Independence (0)
    b. Respect for elders (1)
    c. Neither (.5)
    d. Both (.5)
    e. I don't know (.5)
2. 
    a. Obedience (1)
    b. Self-reliance (0)
    c. Neither (.5)
    d. Both (.5)
    e. I don't know (.5)
3. 
    a. Curiosity (0)
    b. Good manners (1)
    c. Neither (.5)
    d. Both (.5)
    e. I don't know (.5)
4. 
    a. Considerate (0)
    b. Well-behaved (1)
    c. Neither (.5)
    d. Both (.5)
    e. I don't know (.5)

*Cronbach's alpha* = .772

**Contact with Transgender people**
Have you ever had any friends, relatives, or close acquaintances who identify as transgender or as a gender other than male or female?

- Yes (1)
- No (2)

**Tolerance-Intolerance of Ambiguity**

| | Completely Disagree (1) | Somewhat Disagree (2) | Neither agree nor disagree (3) | Somewhat Agree (4) | Completely Agree (5) |
|---|---|---|---|---|---|
| An expert who doesn't come up with a definite answer probably doesn't know too much. | ☐ | ☐ | ☐ | ☐ | ☐ |
| There is really no such thing as a problem that can't be solved. | ☐ | ☐ | ☐ | ☐ | ☐ |
| A good job is one where what is to be done and how it is to be done are always clear. | ☐ | ☐ | ☐ | ☐ | ☐ |
| In the long run it is possible to get more done by tackling small, simple problems rather than large and complicated ones. | ☐ | ☐ | ☐ | ☐ | ☐ |
| What we are used to is always preferable to what is unfamiliar. | ☐ | ☐ | ☐ | ☐ | ☐ |
| A person who leads an even, regular life in which few surprises or unexpected happenings anse, really has a lot to be grateful for. | ☐ | ☐ | ☐ | ☐ | ☐ |
| I like parties where I know most of the people more than ones where all or most of the people are complete strangers. | ☐ | ☐ | ☐ | ☐ | ☐ |
| The sooner we all acquire similar values and ideals the better. | ☐ | ☐ | ☐ | ☐ | ☐ |
| I would like to live in a foreign country for a while. | ☐ | ☐ | ☐ | ☐ | ☐ |
| People who fit their lives to a schedule probably miss most of the joy of living. | ☐ | ☐ | ☐ | ☐ | ☐ |
| It is more fun to tackle a complicated problem than to solve a simple one. | ☐ | ☐ | ☐ | ☐ | ☐ |
| Often the most interesting and stimulating people are | ☐ | ☐ | ☐ | ☐ | ☐ |

| | | | | | |
|---|---|---|---|---|---|
| those who don't mind being different and original. | | | | | |
| People who insist upon a yes or no answer just don't know how complicated things really are. | ☐ | ☐ | ☐ | ☐ | ☐ |
| Many of our most important decisions are based upon insufficient information. | ☐ | ☐ | ☐ | ☐ | ☐ |
| Teachers or supervisors who hand out vague assignments give a chance for one to show initiative and originality. | ☐ | ☐ | ☐ | ☐ | ☐ |
| A good teacher is one who makes you wonder about your way of looking at things. | ☐ | ☐ | ☐ | ☐ | ☐ |

*Cronbach's Alpha* = .766

**Implicit Person Theory**

| | Strongly Agree (1) | Agree (2) | Mostly Agree (3) | Mostly Disagree (4) | Disagree (5) | Strongly Disagree (6) |
|---|---|---|---|---|---|---|
| The kind of person someone is, is something basic about them, and it can't be changed very much. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| People can do things differently, but the important parts of who they are can't really be changed. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Everyone is a certain kind of person, and there is not much that they can do to really change that. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

*Cronbach's alpha* = .970

**Strength of Regulatory Focus**

| | Not at all True for Me (1) | | | | | | | | Very True for Me (9) |
|---|---|---|---|---|---|---|---|---|---|
| In general, I am focused on preventing negative events in my life. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| I am anxious that I will fall short of my responsibilities and obligations. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| I frequently imagine how I will achieve my hopes and aspirations. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| I often think about the person I am afraid I might become in the future. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| I often think about the person I would ideally like to be in the future. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| I typically focus on the success I hope to achieve in the future. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| I often imagine myself experiencing bad things that I fear might happen to me. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| I frequently think about how I can prevent failures in my life. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| I am more oriented toward preventing losses than I am toward achieving gains. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| I see myself as someone who is primarily striving to reach my "ideal self "—to fulfill my hopes, wishes, and aspirations. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| I see myself as someone who is primarily striving to become the self I "ought" to be—to fulfill my duties, responsibilities, and obligations. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| In general, I am focused on achieving positive outcomes in my life. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| I often imagine myself experiencing good things that I hope will happen to me. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Overall, I am more oriented toward achieving success than preventing failure. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

*Cronbach's Alpha* = .881

**Demographics**

Which categories describe you? (Check all that apply)

- ❏ American Indian or Alaskan Native
- ❏ Asian
- ❏ Black or African-American
- ❏ Middle Eastern or North African
- ❏ Native Hawaiian or other Pacific Islander
- ❏ Hispanic or Latino
- ❏ White
- ❏ Other (please specify): _________

How would you describe your sexual orientation?

- o Heterosexual
- o Homosexual
- o Polysexual
- o Something else

What is your gender?

- o Male
- o Female
- o Androgynous
- o Something else

Do you identify as transgender?

- o Yes
- o No

The items below are for background information, please respond using the scale provided.

**Political Orientation**

| | Extremely Conservative<br>1 | 2 | 3 | 4 | 5 | 6 | Extremely Liberal<br>7 |
|---|---|---|---|---|---|---|---|
| How would you describe your political orientation? | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

**Financial Status**

| | Economically Disadvantaged<br>1 | 2 | 3 | 4 | 5 | 6 | Economically Well Off<br>7 |
|---|---|---|---|---|---|---|---|
| How would you describe your household's financial situation | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

## Scale Descriptive Statistics

| Measure | Mean | Std. Deviation |
|---|---|---|
| Transphobia | 31.732 | 12.761 |
| Authoritarianism | 1.156 | 1.074 |
| Tolerance-Intolerance of Ambiguity | 42.907 | 8.222 |
| Implicit Person Theory | 11.402 | 4.481 |
| Regulatory Focus | 56.093 | 18.102 |
| Endorse Binary | 2.856 | 1.258 |

This table shows the scale mean and standard deviation for all measures not reported in the body of the paper.

## Scale Correlations

**Correlations**

| | | Transphobia Score | Authoritarianism Score | Tolerance-Intolerance of Ambiguity | Implicit Person Theory | Regulatory Focus | Endorsement of Gender Binary |
|---|---|---|---|---|---|---|---|
| Transphobia Score | Pearson Correlation | 1 | .361** | .439** | -.302** | .078 | .574** |
| | Sig. (2-tailed) | | .000 | .000 | .003 | .449 | .000 |
| | N | 97 | 96 | 97 | 97 | 97 | 97 |
| Authoritarianism Score | Pearson Correlation | .361** | 1 | .376** | -.153 | .021 | .161 |
| | Sig. (2-tailed) | .000 | | .000 | .136 | .839 | .116 |
| | N | 96 | 96 | 96 | 96 | 96 | 96 |
| Tolerance-Intolerance of Ambiguity | Pearson Correlation | .439** | .376** | 1 | -.269** | .179 | .256* |
| | Sig. (2-tailed) | .000 | .000 | | .008 | .079 | .011 |
| | N | 97 | 96 | 97 | 97 | 97 | 97 |
| Implicit Person Theory | Pearson Correlation | -.302** | -.153 | -.269** | 1 | .021 | -.202* |
| | Sig. (2-tailed) | .003 | .136 | .008 | | .838 | .047 |
| | N | 97 | 96 | 97 | 97 | 97 | 97 |
| Regulatory Focus | Pearson Correlation | .078 | .021 | .179 | .021 | 1 | .074 |
| | Sig. (2-tailed) | .449 | .839 | .079 | .838 | | .470 |
| | N | 97 | 96 | 97 | 97 | 97 | 97 |
| Endorsement of Gender Binary | Pearson Correlation | .574** | .161 | .256* | -.202* | .074 | 1 |
| | Sig. (2-tailed) | .000 | .116 | .011 | .047 | .470 | |
| | N | 97 | 96 | 97 | 97 | 97 | 97 |

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).