Gender categorization and stereotypes beyond the binary

Marie Isabelle Weißflog[1] and Lusine Grigoryan[2]

[1]Department of Social Psychology, Ruhr-Universität Bochum, Bochum, Germany

[2]Department of Psychology, University of York, York, UK

Author Note

Corresponding author: Marie Isabelle Weißflog marie.weissflog@rub.de

https://orcid.org/0000-0003-0686-0581

Lusine Grigoryan https://orcid.org/0000-0002-2077-1975

Abstract

Gender categorization and stereotyping can lead to discrimination. Researchers have mostly studied cisgender, gender-conforming individuals as the targets when examining these processes. Less is known about the gender categorization and stereotyping of gender-ambiguous targets. In two factorial survey experiments, we investigated gender categorization and stereotyping based on facial features and behavioral information. We manipulated femininity/masculinity/ambiguity of face, behavior, and occupation. Participants completed a gender categorization task, and stereotype and attitude measures. The findings indicated that face was most influential for categorization: When face was unambiguously masculine or feminine, participants mostly categorized targets as male or female, respectively. In these cases, behavior and job had little influence on categorization. When face was ambiguous, however, this additional information significantly influenced categorization. Participants categorized 10-13% of targets as nonbinary. Women did so more often than men. Nonbinary categorization was more likely for ambiguous faces, and most likely for ambiguous faces combined with ambiguous behavior and ambiguous or feminine occupation. Targets categorized as nonbinary were rated as warmer than male and female targets, but this was inconsistent across studies. Our findings suggest that categorizing gender-ambiguous targets is more complex compared to clearly gendered targets. Primarily relying on face when it appears clearly gendered likely causes categorization errors when encountering TGNC individuals. The ambiguity of both androgynous faces that are complicated to categorize, and seemingly easily categorizable faces perceived as masculine or feminine that do not align with a person's gender can explain perceived threat and negative bias toward gender-nonconforming individuals in some perceivers.

Keywords: gender non-conformity, social categorization, gender stereotypes, gender nonbinary, transgender, stimulus ambiguity

Gender categorization and stereotypes beyond the binary

Social categorization is central to social cognition. It facilitates orientation in complex social environments through simplification, but can also foster a rigid worldview, resulting in stereotypes and discrimination. Since most contemporary societies are structured by gender, the category is readily accessible (Johar et al., 2003; Stroessner et al., 2010), and individuals habitually categorize others by gender (Banaji & Hardin, 1996; Zarate & Smith, 1990). Gender is tied to stereotypes of appearance, behavior, and roles. Negative consequences of gender categorization and stereotyping include discrimination against women such as violence, constraining traditional roles, and backlash against pursuits of equality (Swim & Hyers, 2009), as well as rigid roles that disadvantage men in terms of mental and physical health (O'Neil, 2008). Traditionally, gender categorization and stereotyping research have focused on cisgender individuals as the targets, yet, gender information is often ambiguous in real life, such as the case with transgender and gender-nonconforming (TGNC) individuals, a fact rarely addressed in studies (Bodenhausen & Peery, 2009).

According to the literature that does exist, individuals rely on different information in ambiguous versus unambiguous categorization situations (Pernet & Belin, 2012). In ambiguous categorization, context cues and secondary information are more important than when primary information is clearly gendered (Huart et al., 2005; D. Martin & Macrae, 2007; Pernet & Belin, 2012). Ambiguous targets are also subject to a general "person = male" bias (K. L. Johnson et al., 2012). Given the higher risk of discrimination and violence against gender non-conforming individuals (Grant et al., 2011; Nadal et al., 2016), as well as mental health problems and suicide (Haas et al., 2011), gender categorization and stereotyping research must move beyond binary gender categories in their studies.

Although gender categorization has been widely studied, many questions remain. What factors influence how individuals categorize and subsequently stereotype others? How does categorization work under volatile, uncertain, complex, and ambiguous (VUCA; Bodenhausen & Peery, 2009) circumstances, which are common in everyday categorization? For example, how does the relative importance of different pieces of information change when other pieces of information are clear vs. ambiguous? In two online factorial survey experiments, we examined the effects of face, occupation, and gender expression on gender categorization and subsequent attitudes and stereotypes. Specifically, we investigated how the masculinity, femininity, and ambiguity of the face, occupation, and expression of the target influences gender categorization and judgments of warmth and competence. Further, we examined potential interaction effects between these factors to determine the relative influence of each on categorization depending on clarity vs. ambiguity of the other factors. Additionally, we investigated how participant gender influences categorization, and how participant political orientation, sexism, and gender attitudes influence attitudes towards gender-ambiguous targets. Both studies were preregistered. The protocols are available on aspredicted.org: https://aspredicted.org/GY1_VYM (Study 1), https://aspredicted.org/JBD_JC1 (Study 2).

The following literature review summarizes findings on gender categorization and stereotypes primarily from Western countries (e.g., US, UK, Belgium), though some studies from other countries (e.g., China) indicate that at least basic gender categorization processes function similarly across cultural contexts. Germany, where we conducted the two studies presented in the current paper, shares common gender roles and stereotypes with other Western countries. However, the terms used to refer to gender are often different from those common in English-speaking countries, such as the US or UK. For example, regarding gender identities, besides female and male, the term “diverse” (German: “divers”) was

introduced as an official third gender marker on German documents in 2018. While originally intended for intersex people only, many LGBTQ+ organizations demand that gender diverse people should be able to claim it as well (Hoenes et al., 2019). The term “nonbinary” (“nichtbinär” or “nonbinär”) is also slowly gaining more public awareness in Germany.

Gender Categorization

Social categorization simplifies human interaction by structuring the social environment and making it more predictable. It also complicates human interaction by enabling stereotypes, prejudice, and discrimination. Fit (i.e., how meaningfully members of different categories differ in their actual characteristics, social norms, and roles) and accessibility (i.e., how salient a category is to an individual due to chronic and situational factors) determine which categories an individual uses in a specific situation (Oakes, 1978; Oakes et al., 1991). We apply the category “gender” extremely quickly and often (Banaji & Hardin, 1996; Zarate & Smith, 1990). It is often situationally accessible because it is linked to relevant role expectations reflected in individuals’ behavior, gendered language, etc. (Stroessner et al., 2010). High perceived fit and frequent situational accessibility, along with individual stereotypes, make gender categories chronically accessible (Johar et al., 2003). Several types of gender-stereotyped information influence categorization. Face plays an important role, as well as expression and occupation.

Categorization by Face

Faces are central to social cognition because they help humans recognize and categorize each other. Basic facial gender categorization happens automatically, indicating that humans easily extract gendered information from faces. Zhang et al. (2016) found that gender categorization happens within 195 ms after seeing a face and precedes stereotype activation (at 475 ms). Almost all facial features carry gendered information, especially the jaw, brows, eyes, and chin (Brown & Perrett, 1993). Beyond single features, holistic

impressions also play an important role (Zhao & Hayward, 2010). When deciding if a face is female or male, individuals decide categorically, even when faces are morphed continuously. According to the face-space metaphor theory, individuals compare each face to a prototypical face or an "attractor field" of prototypical features of each category (Campanella et al., 2001), with higher closeness to the prototype indicating higher similarity. In line with this theory, O'Toole et al. (1998) reported easier categorization of caricatural masculine or feminine faces. Based on the literature, we expected that targets with ambiguous faces would be more likely to be categorized as nonbinary compared to targets with masculine features (H1). The preregistered hypothesis was based only on ambiguous vs. masculine contrast. However, to examine how nonbinary categorization is used in both male and female categorization, we tested a model with ambiguous/nonbinary as reference category to get both contrasts of ambiguous/nonbinary vs. masculine/male and feminine/female, respectively.

Categorization by Expression

Gender presentation or expression is another factor in gender categorization. The American Psychological Association (APA) defines gender presentation as "[t]he presentation of an individual, including physical appearance, clothing choice and accessories, and behaviors that express aspects of gender identity or role. Gender expression may or may not conform to a person's gender identity" (American Psychological Association, 2015). Expression thus includes many different characteristics that may cue a perceiver around gender categorization, such as hairstyle, choice of clothing, and use of make-up (Hayfield, 2013; Horn, 2007), as well as sports and leisure activities (Lippa, 2005; Plaza et al., 2017; Taylor, 2003). Descriptive and prescriptive gender stereotypes often include aspects of gender expression, like preferred activities and appearance (Koenig, 2018). Even when individuals do not personally endorse gender stereotypes, they may still use them for categorization when a task requires it. Further, perceivers' own gender role orientation

influences which information they use for categorization: individuals with a stronger gendered self-schema are more likely to interpret expression information in a gendered way (Skitka & Maslach, 1990). We expected an ambiguous vs. masculine gender expression to increase nonbinary vs. male categorization (H2). To get both contrasts between ambiguous/nonbinary and masculine/male as well as feminine/female, respectively, we used ambiguous/nonbinary as the reference category.

Categorization by Occupation

Occupations are often associated with gender (Glick et al., 1995; Karylowski et al., 2001; Oswald, 2003; White et al., 1989; White & White, 2006), which is a primary component of occupational stereotypes besides prestige (Glick et al., 1995). For example, secretary and kindergarten teacher are stereotypically feminine occupations, while carpenter and engineer are stereotypically masculine. These stereotypes include the perceived distribution of women and men within occupational groups, as well as beliefs about occupation-specific competence differences by gender (Canessa-Pollard et al., 2022; Levy et al., 2000). Occupational gender stereotypes change over time, with some becoming less and others becoming more stereotyped (White et al., 1989). Even in occupations rated less gendered than they used to be (e.g., accountant), implicit judgments can still show stronger gender bias (White & White, 2006). Therefore, the link between gender and occupation is still highly relevant.

We expected ambiguous vs. masculine job to increase nonbinary vs. male categorization (H3). To get both contrasts between ambiguous/nonbinary and masculine/male as well as feminine/female, respectively, we used ambiguous/nonbinary as reference category.

Ambiguous Categorization

Research on social categorization and its consequences has often focused on one clear category. However, in everyday life, categorization occurs under more complicated conditions. Bodenhausen and Peery (2009) highlighted how considering conditions of volatility, uncertainty, complexity, and ambiguity (VUCA) helps better understand real-life categorization. VUCA factors often appear in combination and can be hard to disentangle, but we will focus on ambiguity here. There is a male bias in ambiguous cases wherein people tend to assume a target person is a man unless otherwise specified (Hamilton, 1991). For example, participants categorize ambiguous body shapes as male more readily than as female (K. L. Johnson et al., 2012). Individuals use some information for categorization only when more commonly used information is ambiguous. For example, people usually rely on timbre to categorize voices, but when timbre is ambiguous, they also consider pitch information (Pernet & Belin, 2012). This might apply to other factors in categorization as well, but as far as we are aware, has not been applied to face yet.

In face categorization, gendered context cues like hairstyles or names can bias categorization in a more feminine or masculine direction (Huart et al., 2005; Macrae & Martin, 2007), especially but not exclusively in ambiguous faces. However, this effect might be particularly strong when face is ambiguous and therefore provides less reliable information. Context can also shape gender perception and categorization by providing a standard of comparison for femininity or masculinity. Relatively gender-neutral faces, for example, are perceived as more masculine and more likely to be male when presented after more feminine faces, and vice versa (A. E. Martin, 2022). We explored how these features interacted in predicting nonbinary categorization. We were especially interested in the influence of expressions and occupation information when face was clearly ambiguous vs.

clearly feminine or masculine. We also explored how frequency of nonbinary category use was related to participant characteristics and attitudes.

Gender Stereotypes

One framework to understand stereotypes is the Stereotype Content Model (SCM; Fiske et al., 2002), which proposes two fundamental dimensions of stereotype content that guide social perception: warmth and competence (also called communion and agency; Abele & Wojciszke, 2014). Warmth refers to traits and behaviors relevant to social functioning, like kindness and trustworthiness. Competence refers to traits and behaviors relevant to goal-achievement, like efficacy and confidence. Past research on gender stereotypes has mostly focused on cisgender women and men. Both men and women perceive women as warmer than men and men as more competent than women in some, but not all domains (Hentschel et al., 2019), while other research has found equality in competence perceptions or even a female advantage (Eagly et al., 2020). Some research has also examined transgender stereotypes (Gazzola & Morrison, 2014; Howansky et al., 2019), demonstrating that stereotypes about transgender men and women overlap more with their assigned gender at birth than those for cisgender men with those for cisgender women (Howansky et al., 2019). There are also specific transgender stereotypes related to perceived social outcast status which the SCM may not capture (Gazzola & Morrison, 2014).

Research on nonbinary stereotypes is very limited. In gender ambiguous cases, individuals assume that physically androgynous individuals have less gender-typed behaviors, personality traits, roles, and occupations than clearly gendered targets (Madson, 2000). Stereotype differences between ambiguous targets categorized as female vs. male are smaller than usual gender stereotype differences (Madson, 2000). Individuals may be less sure about categorization of ambiguous targets compared to clearly gendered targets, thus relying more on individuating information than stereotypes to make inferences about them. In

the current study, we explored how assumed gender is related to stereotypes of warmth and competence, as well as warmth towards targets.

Overview of Current Study

We conducted two studies using a factorial survey experimental design to test our research questions. Study 1 was a first test of the design in a convenience sample with materials that had only been partially validated beforehand. Study 2 was mostly a replication of the design from Study 1 in a more representative sample with all materials pre-tested.

Method – Study 1

All materials, data, and code for both studies are available on OSF: https://osf.io/encr3/?view_only=e38b15b54cbb4db49fd7f6bcb5a7712a

Participants and Procedure

The smallest effect of interest was the indirect effect of the different features on stereotypes, mediated by categorization. We do not report this analysis in the article, as we found that stereotypes did not differ by the perceived gender category, therefore the assumptions for testing the mediation model were not met. Expecting small to medium sized mediated effects, we aimed for at least 148-162 participants to detect an effect of that size (Fritz & MacKinnon, 2007), and 50 participants per condition (Maas & Hox, 2005) to avoid biased standard error estimates in multilevel models. Thus, we aimed for 200 participants to compensate for possible missing data. A total of 239 participants started the survey and 204 (85.4%) completed it. For our analyses, we used all available data from the vignette section of the survey, even when participants did not complete the entire survey. The sample size that was available for each analysis is specified in the respective results table. Recruitment took place between March-May 2020. The study was advertised on an online course credit platform for psychology students Ruhr University Bochum and in a corresponding Facebook group. The study was also advertised on SurveyCircle (2020), offering participants a code to

promote their own study. Participants self-selected into the study. The only inclusion criterion was being at least 18 years old. The study was approved by the ethics review board of Ruhr University Bochum.

The average age was 28.5 years (*SD* = 10.5). Most participants identified as female (78.5%, *n* = 157), 21.5% as male (n = 43), and none as nonbinary. Most (74.5%, n = 149) identified as German. The majority (56%, n = 112) was not connected to any religion, 22% (44) were protestant, and 15% (30) Roman Catholic. Regardless of religion, participants were on average only slightly religious (*M* = 3.4, *SD* = 2.4, on a 10-point scale from "*not religious at all*" to "*very religious*"). Most participants had completed high school (42.5%, n = 85), 26.5% (53) had a Bachelor's degree, and 13.5% (27) had a Master's degree. For employment, 27.1% (54) were not working (including students without side jobs), 21.1% (42) were employed full-time, and 38.7% (36) were employed part-time. On a scale from 1-4 indicating how well participants could live with their household income, they were on average able to get by (*M* = 3.3, *SD* = 0.7). Average political orientation was moderately leftist (*M* = 4.6, *SD* = 1.6, on an 11-point scale from "*extreme left*" to "*extreme right*")'.

Data were collected on the online survey platform Qualtrics (2020) via an anonymous link. The study consisted of a factorial survey experiment and questionnaires. In the factorial survey, we manipulated face, occupation, and expression across three levels: feminine, ambiguous, and masculine (see Table 1). Vignettes were split into three blocks of nine vignettes to prevent participant fatigue. Blocking was done with the D-efficiency algorithm in SAS (SAS Institute Inc., 2015), which ensures that independent variables remain orthogonal and all levels are presented in a balanced way across blocks. All participants gave informed consent. They were then given instructions on how to answer the vignette questions. They were randomly assigned to one of three vignette blocks, each vignette followed by a gender categorization task and assessments of warmth and competence as well as warmth towards

targets. Participants then completed a demographic questionnaire, followed by a set of exploratory measures for gender identification, ambivalent sexism, modern sexism, and gender role. Finally, participants had to confirm that their data should be used for analysis and were then given course credit or SurveyCircle code.

Materials

Vignette composition. Vignettes consisted of three components: a picture of a face followed by a description of the target's occupation and expression.

Face. Faces were selected from the Chicago Face Database (Ma et al., 2015) which includes norming data for masculinity, femininity, perceived gender, attractiveness, and emotional expression. We chose faces with a neutral expression since emotional expression may impact categorization (Smith et al., 2015; Zhang et al., 2019). We only included White faces, because including race as another factor in the analysis would have exceeded the scope of the current study. However, this design limits *generalizability.* Masculine faces were selected by masculinity > 5 and femininity < 2 on a 7-point scale, the strictest criterion that still yielded exactly three White masculine faces. Feminine faces were selected by femininity > 5 and masculinity < 2, yielding 11 images. We chose three with no or minimal make-up. Ambiguous faces were morphed from masculine and feminine faces in InterFace (Kramer et al., 2017). Each masculine and feminine face was used to create one morphed face, yielding three ambiguous images. All images were cropped into ovals to remove hair cues which may have influenced categorization. Using three faces per gender condition helps limit stimulus-specific effects.

Gender Expression. We chose clothing and make-up for gender expression which are reliably part of stereotypes (Hayfield, 2013; Horn, 2007; Koenig, 2018). Descriptions consisted of a short sentence reading "The person [always/sometimes/never] wears make-up and likes [dressing up/sometimes likes dressing up/does not like dressing up]."

Occupation. Occupation descriptions consisted of a short sentence: "The person works as [.....]" followed by an occupation. Occupations were selected from Oswald (2003), Sinclair and Carlsson (2013), and White et al. (1989), in which occupations had been rated for perceived femininity/masculinity. We selected mostly feminine, mostly masculine, and ambiguous occupations accordingly. We selected occupations with similar social status to eliminate status as a potential confounding variable (Glick et al., 1995; Oswald, 2003). For that purpose, occupations were coded according to their occupational group in the International Standard Classification of Occupations-08 (ISCO-08; International Labour Organization, 2012). We calculated status for each occupation using the Standard International Occupational Prestige Scale-08 (SIOPS-08; Ganzeboom & Treiman, 2003) and the International Socio-Economic Index of Occupational Status-08 (ISEI-08; Ganzeboom, 2010) using the R package ISCO08ConveRsions (Schwitter, 2019). For feminine occupations, we finally chose beautician, dental hygienist, and kindergarten teacher. For ambiguous occupations, we chose restaurant manager, travel agent, and photographer. For masculine occupations, we chose firefighter, athletic trainer, and police officer.

Gender categorization. After each vignette, participants were asked "What do you think is the gender of this person?" Response options were "male," "female," and "diverse/nonbinary." We chose "diverse/non-binary" for a German sample since the term "diverse" has gained public awareness in Germany after its introduction as an official third gender marker. We used "nonbinary" as well since it is a more descriptive umbrella term, albeit less known in Germany. Figure 1 shows an example of the stimuli and gender categorization task.

Warmth and competence. We used four items available in German from Abele et al. (2016) to measure warmth and competence. Warmth items were scored on a response scale from 1 ("*not at all friendly*") to 5 ("*very friendly*") and 1 ("*not trustworthy*") to 5 ("*very*

trustworthy).” Competence items were scored on a scale from 1 (“*little competent*”) to 5 (“*very competent*”) and 1 (“*has no leadership abilities at all*”) to 5 (“*has leadership qualities*”) using the same 5-point scale. Mean warmth ratings were 3.40 (*SD* = 0.90) with Cronbach’s alpha = 0.79. Mean competence ratings were 3.30 (*SD* = 0.80) with Cronbach’s alpha = 0.60. Internal consistency was lower than in the original study, perhaps due to the smaller number of items.

Warmth towards targets. Warmth towards targets was measured with a feeling thermometer, with zero degrees meaning very cold and ten degrees meaning very warm. Average warmth towards targets was 4.70 (*SD* = 2.10).

Sexist Attitudes. Benevolent and hostile sexist attitudes were measured with the German version of the Ambivalent Sexism Inventory (ASI; Glick & Fiske, 1996, German version by Eckes and Six-Materna, 1999). Mean Benevolent Sexism score was 3.0 on a response scale from 1 (“*do not agree at all”*) to 6 (*“completely agree”*), *SD* = 0.9, Cronbach’s alpha = .85. Mean Hostile Sexism was 2.7, *SD* = 1.0, Cronbach’s alpha = .93. Modern Sexism was measured with the German Modern Sexism Scale (MSS; Eckes & Six-Materna, 1998); *M* = 3.0, *SD* = 0.8, on a scale from 1 (“*do not agree at all”*) to 6 (*“completely agree”*). Cronbach’s alpha was .89.

Gender Identification. Gender identification was measured with four items adapted from Becker and Wagner (2009) to refer to gender identification in general instead of identification with womanhood specifically: *“I identify with other people of my gender”, “I feel strong ties to other people of my gender”, “Overall, my gender is an important part of my self-image”, and “Belonging to my gender is important to me”*. Mean gender identification was 4.20, *SD* = 1.00, on a scale of 1 (*“strongly disagree”*)-6 (*“strongly agree”*). Cronbach’s alpha was .78. Gender role orientation was measured with the revised German Bem Sex-Role Inventory (BSRI-R; Troche & Rammsayer, 2011). Cronbach’s alpha

was .89 for masculinity items and .90 for femininity items. A total of 31.8% (63) of participants identified as undifferentiated, 29.3% (58) as androgynous, 20.2% (40) as feminine, and 18.7% (37) as masculine. We included two attention check items in the BSRI-R and the ASI.

Demographics. Demographic questions included age, gender (male/female/diverse), migration background, cultural background, highest education, and current employment. Political orientation was measured on an 11-point scale from "*extremely left-wing*" to "*extremely right-wing*." Religiosity was polled using two items, one asking if/which religious community participants felt connected to and the other measuring religiosity independent of affiliation on a scale from 1 = "*not at all religious*" to 10 = "*very religious*."

Data Analysis

Analyses were conducted with R (R Core Team, 2013)/RStudio (RStudio Team, 2020) and MPlus (Muthén & Muthén, 2010). Two participants were excluded because they indicated their data should not be used as they did not answer seriously or had already taken part before. Four participants were also excluded because they failed both attention checks. These exclusions deviate from the preregistration where we stated there would be no post hoc exclusion criteria. Effective sample size after exclusions was 233. Missing data summary with the R package naniar (Tierney et al., 2020) yielded 3.6% missing data. Since missing data percentages below 5% (Schafer, 1999) or even 10% (Bennett, 2001) are unlikely to bias analyses, we took no further steps to analyze missing data patterns. We did not exclude any outliers since the statistical procedures used were relatively robust to outliers.

As the data were nested (vignettes nested within respondents), we used cluster-robust standard errors or multilevel models for all analyses involving vignette data. We tested influence of features on categorization with multinomial multilevel regression using MPlus. In addition to the preregistered analysis where we used "male" and "masculine" as the

reference categories, we also tested the same models using "nonbinary" and "ambiguous" as reference categories, to see how the nonbinary category is differentiated from both male and female categories. We tested the link between gender categorization and stereotypes and warmth towards targets with multilevel multinomial logistic regression, and the link between participant characteristics and the frequency of nonbinary category use with Poisson and negative binomial regression for count data.

Results – Study 1

Use of Different Gender Categories

Table 2 shows how often participants selected each gender category during the categorization task. They used the male category half of the time(50%, n = 917), followed by the female category (40%, n = 734) and then the nonbinary category (10%, n = 187).

Table 3 shows how frequently participants used the nonbinary category depending on gender, age, political views, and religiosity (Step 1), and modern sexism, ambivalent sexism, gender identification, and sex role (Step 2). Women used the nonbinary category more than men (b = 0.72, SE = 0.29, p = .015). However, with sexism and gender identification/role considered, this effect was no longer significant. More religious participants used the nonbinary category less (b = -0.13, SE = 0.05, p = .005). No other factor influenced frequency of nonbinary category use.

Effects of face, gender expression, and occupation on gender categorization

Table 4 shows the influence of face, occupation, and expression on nonbinary vs. male and female categorization, respectively. Participants relied mostly on facial stimuli to categorize targets. Participants were much more likely to categorize a target as nonbinary vs. male when face was ambiguous vs. masculine (OR = 9.49, p <.001), supporting H1. Ambiguous vs. masculine occupation had no significant effect on nonbinary vs. male categorization (OR = 0.73, p = .085), not supporting H2. Still, participants were more likely

to categorize a target as nonbinary vs. male when expression was ambiguous vs. masculine (*OR* = 2.16, *p* <.001), supporting H3.However, the influence of behavior and job on categorization changed with face ambiguity. Figures 2, 3, and 4 illustrate that expression and occupation had essentially no influence on categorization when face was clearly masculine or feminine. However, when face was ambiguous, this additional information was considered and influenced categorization results. Surprisingly, masculine expression and occupation both increased the likelihood of female categorization, and vice versa for feminine expression and occupation.

Figure 2 shows the percentage of targets that were categorized as nonbinary for different combinations of face, job, and behavior. There is, again, a pattern that nonbinary categorization is more likely when face is ambiguous, and this varies depending on expression and occupation. In contrast, when face is clearly masculine or feminine, nonbinary categorization is unlikely, with little difference between different expression-occupation configurations. Table 5 contains the results of the underlying interaction model. We also conducted an exploratory multiple correspondence analysis (MCA) of features and categorization outcomes (see Figure 3). It showed that the two most important dimensions on which the data are differentiated appear to be a distinction between masculine faces categorized as male and feminine faces categorized as female (Dimension 1), as well as between those two sets of faces and ambiguous faces that are more often categorized as nonbinary (Dimension 2). Faces and categorization outcomes are far away from the origin, which means they strongly differentiate between observations. Interestingly, this is somewhat less pronounced for ambiguous faces and nonbinary categorization, hinting at more complex categorization processes at work that may be influenced by additional factors that are not captured well within these first two dimensions.

We observed a male bias in ambiguous face categorization: 54.5% (333) of ambiguous faces were categorized as male, and the remainder was split between female (21.9%, n = 134) and nonbinary (23.6%, n = 144) categories.

Stereotypes and warmth towards targets

Table 6 shows the association between gender categorization and warmth and competence stereotypes as well as warmth towards targets. Surprisingly, gender categorization predicted neither endorsement of stereotype content nor warmth towards targets. We also tested this model with only the contrast between female vs. male categorization instead of all three gender categories, still finding no effect except that targets categorized as female vs. male were liked more. Face, behavior, and job did not directly predict warmth towards targets, either (see Table 7).

Discussion – Study 1

Participants used the nonbinary category for 10% (187) of the categorization stimuli. 77% (144) of faces categorized as nonbinary were ambiguous faces, 15% (28) masculine faces, and 8% (15) feminine faces. Given the study design was balanced with masculine, feminine, and ambiguous characteristics presented equally often, this finding may indicate hesitancy to use the nonbinary gender category. However, given the estimated prevalence of TGNC individuals across societies is around 1-2% (Goodman et al., 2019; Spizzirri et al., 2021), suggesting that participants may have adjusted their judgments by the likelihood of someone identifying as nonbinary. We also found that more religious individuals were more hesitant to use the nonbinary category, which fits previous findings that show a link between religiosity or religious fundamentalism and prejudice against transgender people as well as a preference for binary gender conceptions (Adams et al., 2016; Broussard et al., 2018; Norton & Herek, 2013; Worthen et al., 2017).

The results of the multilevel multinomial model clearly indicate that facial features were most important in gender categorization. When face was unambiguously masculine or feminine, participants categorized the target as, respectively, male or female in almost 100% of cases, irrespective of occupation and expression. However, when face was ambiguous, participants' judgments were influenced by targets' expression and occupation as well. Targets with ambiguous faces were generally more likely to be categorized as nonbinary than targets with clearly masculine or feminine faces. The exploratory MCA showed that targets were primarily differentiated by face, on two dimensions contrasting feminine/female vs. masculine/male, and ambiguous/nonbinary vs. unambiguous, respectively.

We also observed clear evidence of male bias (Hamilton, 1991). A full 50% of all targets were categorized as male despite only a third of all stimuli being masculine. This over-use of the male category was explained by targets with ambiguous faces being categorized as male twice as often as female or nonbinary. This aligns with previous findings on body shape categorization where participants categorized ambiguous shapes as male more readily than as female (K. L. Johnson et al., 2012).

This study has several limitations that should be considered. First, the sample was predominantly young (M = 28.5 years, SD = 10.5), female (78.5% compared to 21.5% male), and highly educated. It included fewer participants with a migration background than the German average (18.9% vs. German average: 25.5%; Statistisches Bundesamt Destatis, 2019). Importantly, no participants were nonbinary/diverse. Second, not all stimuli were pretested. Norming data were available for masculine and feminine faces, but not for ambiguous faces, so we cannot be sure that participants indeed perceived these faces as ambiguous. Additionally, facial stimuli were not controlled for attractiveness, even though participants may have perceived ambiguous targets as less attractive (Madson, 2000). Given the scarce research on gender expression cues in categorization, selecting stimulus material

for this variable was difficult. Our stimuli included make-up and clothing, which covers only one of many aspects of gender expression. Make-up and clothing may also influence gender categorization differently which would not be evident from the combined wording. Always dressing up and wearing make-up may not represent the common stereotypes about women's appearance in Germany and in the present sample. We addressed these limitations in Study 2.

Method – Study 2

In Study 2, we aimed to replicate Study 1 with a more representative sample. We improved the design by pretesting the vignette components to make sure that faces were matched by attractiveness, and that all components were perceived in alignment with intended gender stereotypes.

Pretest of Experimental Stimuli

Participants and Procedure

Fifty participants were recruited in January 2021 via Prolific (2021). They were paid €2.20 for participation. Participants had to be German nationals, 18 years or older, currently residing in Germany. The study was approved by the ethics review board of Ruhr University Bochum. Participants were presented with 36 facial stimuli and asked to rate each on a 7-point scale from 1 ("*not at all"*) to 7 ("*very"*) on masculinity, femininity, androgyny, and attractiveness. We used the term "androgynous" in the item because it is more widely used and understood than "gender-ambiguous." Afterwards, participants were presented with 43 occupation titles, followed by 41 activities or behaviors, rating each on the same 7-point scale for masculinity, femininity, and androgyny.

Stimuli

Faces were selected using the same criteria as in Study 1. For a greater selection of ambiguous faces to choose from, we morphed each feminine face with two of the masculine

faces instead of only one. Occupations were drawn from the same item pool as in Study 1. Expression included the behaviors of make-up and clothing from Study 1, as well as hobbies and activities with different gendered associations drawn from past research (Lippa, 2005; Plaza et al., 2017; Taylor, 2003).

Results

We calculated mean masculinity, femininity, androgyny, and – if applicable – attractiveness scores for each item, as well as differences between the mean scores. We conducted linear regression analyses to check if gender ratings or difference scores predicted facial attractiveness. We selected masculine and feminine stimuli by filtering for masculinity or femininity respectively as well as difference scores that were at least one SD above mean. We selected androgynous stimuli by filtering for androgyny scores at least one SD above mean for occupations and behaviors, while filtering for androgyny and difference scores at least ¼ SD above mean for faces.

Face.

Mean androgyny of faces was 3.4 (SD = 1.0), mean femininity was 3.8 (SD = 1.5), and mean masculinity was 4.0 (SD = 1.7). Mean attractiveness was 3.4 (SD = 0.7).

Three male faces were perceived as masculine. These were the same as the three masculine faces chosen in Study 1, which were thus confirmed to be perceived as masculine. Eight female faces were perceived as feminine, including only two out of three of the feminine faces chosen in Study 1. Ten faces were perceived as androgynous, nine of which were morphed faces, and one was a female face. None of the morphed faces chosen in Study 1 were among the androgynous faces according to the pretest. Face filtering results and their respective item ratings are shown in Table 6 in Online Resource 1. Attractiveness was negatively associated with masculinity (b = -.16, SE = .06, p = .02) and androgyny (b = -.24, SE = .11, p = .03), and positively associated with femininity-androgyny difference (b = .12,

SE = .06, *p* = .05), suggesting that attractiveness needs to be considered when selecting the faces.

Occupation. Mean masculinityof occupations was 4.8 (*SD* = 1.1), mean femininity 4.4 (*SD* = 1.2), and mean androgyny 3.8 (*SD* = 0.7). Six occupations were perceived as masculine – firefighter, construction worker, taxi driver, garbage collector, forester, and marine officer. This included only one masculine occupation used in Study 1, firefighter, but not athletic trainer or police officer. Six occupations were perceived as feminine – beautician and kindergarten teacher, which were also included as feminine occupations in Study 1, nurse, hairdresser, housekeeper, and elementary teacher. Dental hygienist,which was included in Study 1, was not included in this selection. Seven occupations were perceived as androgynous, - photographer, which was also included as an androgynous occupation in Study 1, architect, psychologist, artist, musician, reporter, and university teacher. The other androgynous occupations from Study 1, travel agent or restaurant manager, were not included in the selection. Occupation filtering results and their respective item ratings are shown in Table 7 in Online Resource 1.

Expression. Mean masculinity of gender expression behaviors was 4.5 (*SD* = 1.5), mean femininity 4.6 (*SD* = 1.4), and mean androgyny 3.9 (*SD* = 0.8). Nine expression behaviors were perceived as masculine – never wearing make-up, which was also used in Study 1, rugby, ice hockey, bodybuilding, car racing, boxing, home improvement, strength training, and dart. Never dressing up, which was also part of the masculine items in Study 1, was not included in the selection. Four expression behaviors were perceived as feminine – embroidery, cheerleading, riding, and always wearing make-up, which was also used in Study 1. However, always dressing up which we also used as a feminine behavior in Study 1 was not included in the selection. Eight behaviors were perceived as androgynous, including none of the items from Study 1. Items perceived as androgynous in the pretest were always

dressing up, sometimes dressing up, swimming, skiing, social drinking, cooking, acting, and running. Expression behavior filtering results and their respective item ratings are shown in Table 8 in Online Resource 1.

Discussion

Findings from the pretest indicated that not all items from Study 1 were perceived as intended, highlighting the importance of a follow-up study with pretested materials. Further, since attractiveness ratings varied by gender ratings, faces should be matched by attractiveness across gender conditions to prevent distortion by attractiveness stereotypes.

Main Study

Participants and Procedure

Power considerations were the same as in Study 1, so we again aimed for 200 participants. Participant recruitment took place in February 2021 via Prolific. Participants were paid €2.20. Participants had to be German nationals, 18 years or older, and currently residing in Germany. The study was approved by the ethics review board of Ruhr University Bochum. A total of 201 participants started the survey and 200 (99.5%) completed it. For our analyses, we used all available data from the vignette section of the survey, even when participants did not complete the entire survey. The sample size that was available for each analysis is specified in the respective results table.

Participants were 29 years old on average (*SD* = 8.4). 2% (4) were nonbinary/diverse, 34.5% (69) female, and 63.5% (127) male. A total of 76% (152) had no migration background and 72.5% (145) described their cultural background as only German. For education, 31.5% (63) had completed high school, 29% (58) had a Bachelor's degree, and 18.5% (37) had a Master's degree. For employment, 37.5% (75) were employed full-time, 28.0% (56) not currently working (including students without side jobs), 26.0% (29)

employed part-time. Mean subjective income indicated participants were, on average, able to get by on their household income (M = 3.1, SD = 0.7). Average political orientation was moderately leftist (M = 4.5, SD = 1.8, on an 11-point scale from "*extremely left*" to "*extremely right*"). The majority (63.5%, n = 127) were not connected to any religion, 14% (28) were Protestant, and 13.5% (27) Roman Catholic. Participants were on average only slightly religious (M = 2.9, SD = 2.3, on a 10-point scale from "*not religious at all*" to "*very religious*").

Data were collected via online survey platform Qualtrics (2021) with an anonymous link. The study design was the same as in Study 1. Vignette blocking was done with the fac.design function from R package DoE.base (Grömping, 2018). The questionnaire structure was mostly the same as in Study 1. In the exploratory questionnaires, the BSRI and ASI were replaced by measures of openness towards nonbinary gender and gender essentialism.

Materials

Vignette composition was the same as in Study 1, but we adjusted the content of each factor according to the pretest (see Table 8). Gender categorization was measured with the same item as in Study 1. Warmth and competence were measured with one item each: "*friendly*" for warmth (*M* = 3.3, *SD* = 0.9) and "*competent*" for competence (*M* = 3.4, *SD* = 0.8). Warmth towards targets was measured with a feeling thermometer as in Study 1 (*M* = 4.6, *SD* = 2.2).

Individual Difference Measures. Modern Sexism was again measured with the German Modern Sexism Scale (Eckes & Six-Materna, 1998), with Cronbach's α = .91, *M* = 3.1, *SD* = 1.0. Gender identification was measured with the same items as in Study 1, with Cronbach's α = .81, *M* = 3.9, *SD* = 1.1. Attitude towards nonbinary gender was measured with the Openness towards Non-Binary Gender (ONBG) scale (Molin et al., 2020), on a scale ranging from 1 (*"do not agree at all"*) to 7 (*"completely agree"*), with Cronbach's α = .95, *M* = 4.60, *SD* = 1.50. Example item: *"There are more than two gender categories."* Gender essentialism was measured with the Gender Essentialism Scale (GES; Skewes et al., 2018), on a scale ranging from 1 (*"do not agree at all"*) to 7 (*"completely agree"*), with Cronbach's α = .93, *M* = 3.50, *SD* = 1.10. Example item: **"***Male and female brains probably work in very different ways."*The demographic questionnaire was the same as in Study 1.

Analysis

Analysis software was the same as in Study 1. No participants were excluded. Missing data summary yielded 0.07% missing data; therefore, we took no further steps to analyze missing data patterns. We used the same statistical procedures as in Study 1 for the corresponding hypotheses.

Results – Study 2

Use of Different Gender Categories

Table 9 shows how many targets were placed in each category. Participants used the female category most (45%, n = 801), followed by the male category (42%, n = 758). The nonbinary category was used in 13% (240) of the cases. Women used the male category most, followed by the female category. For men, it was the opposite. Diverse or nonbinary participants used the female category most, followed by the male category. Table 10 shows how frequently participants used the nonbinary category depending on gender, age, political views, and religiosity (Step 1), as well as modern sexism, gender essentialism, gender identification, and openness towards nonbinary gender (Step 2). Women used the nonbinary category more often than men (p = 0.33, SE = 0.16, p = .038), but the effect was no longer significant in Step 2 of the model. No other variable predicted frequency of nonbinary categorization.

Effects of face, gender expression, and occupation on gender categorization

Table 11 shows the influence of face, occupation, and expression on nonbinary vs. male and female categorization, respectively. Face had the strongest influence on categorization. Participants were much more likely to categorize a target as nonbinary vs. male when face was ambiguous vs. masculine (OR = 464.05, p <.001), supporting H1. As expected, when expression was ambiguous vs. masculine, participants categorized targets as nonbinary vs. male more often (OR = 1.52, p = .029), supporting H2 Ambiguous vs. masculine occupation had no effect on nonbinary vs. male categorization, so H3 was not supported.

Additional analysis with female and feminine as reference categories supports the dominant influence of face. Participants were much more likely to categorize targets as nonbinary vs. female when face was ambiguous vs. feminine (OR = 76.71, p <.001).

Occupation, again, had no effect. As expected, when expression was ambiguous vs. feminine, targets were more likely to be categorized as nonbinary vs. female (*OR* = 1.55, *p* = .030).

Consistent with Study 1, face was a defining factor in gender categorization. When it was clearly feminine or masculine, participants categorized targets as female or male respectively almost always, regardless of their expression or occupation. However, when face was ambiguous, participants relied more on the additional information from occupation and expression to categorize the target (see Figure 4). Unlike in Study 1, we did not observe male bias in the categorization of targets with ambiguous faces: they were categorized as male, female, or nonbinary with similar frequencies. This may be due the fact that the ambiguous faces we used in Study 1 were, on average, perceived as more masculine in the pre-test in Study 2 than the ambiguous faces we then chose for Study 2 (see Online Resource 1, Table 1). Masculine expression increased the likelihood of male categorization and feminine expression the likelihood of female categorization when face was ambiguous. As in Study 1, we found that masculine occupation increased the likelihood of female categorization when face was ambiguous. Nonbinary categorization was most likely when an ambiguous face was combined with ambiguous or masculine expression and/or feminine occupation.

Figure 4 shows the percentage of targets categorized as nonbinary for different combinations of face, occupation, and expression. Similar to Study 1, we see a pattern that nonbinary categorization is more likely when face is ambiguous, and how likely exactly varies depending on occupation and expression. When face is clearly masculine or feminine, nonbinary categorization is again very unlikely, with little difference between different behavior-job-configurations. Table 12 contains the results of the underlying interaction model.

Again, we conducted an MCA of features and categorization outcomes (see Figure 5). Like in Study 1, the two most important dimensions differentiated ambiguous faces categorized as nonbinary from the rest (Dimension 1), as well as masculine from feminine faces, each categorized accordingly (Dimension 2). Again, faces and categorization strongly differentiate between observations, which is also again less pronounced for ambiguous faces. Like in Study 1, this may point to more complex categorization processes in ambiguous face categorization.

Stereotypes and Warmth towards targets

Table 13 shows the relationship between gender categorization and warmth and competence stereotypes as well as warmth towards targets. Targets categorized as nonbinary were rated warmer than targets categorized as male or female. Besides that, gender categorization surprisingly predicted neither stereotypes nor warmth towards targets. Again, we also tested this model with only the contrast between female vs. male categorization, finding no stereotype differences between the groups. Contrary to Study 1, face, expression, and occupation did in some cases predict stereotypes directly (see Online Resource 1, Table 5). Ambiguous faces were rated warmer than both masculine and feminine faces. Both feminine and masculine jobs were rated warmer than ambiguous occupations, and masculine occupations were also rated more competent. Expression had no direct influence on warmth and competence stereotypes. Regarding warmth towards targets, participants rated ambiguous faces more positively compared to masculine faces, and feminine and masculine occupations more positively than ambiguous occupations. Expression did not influence warmth towards targets directly.

Discussion – Study 2

Participants used the nonbinary category in 13% of the cases, similar to the pattern observed in Study 1. Women selected the nonbinary category more often than men, but when

modern sexism, gender essentialism, gender identification, and openness towards nonbinary gender were included in the model, the effect was no longer significant. None of the other participant characteristics influenced nonbinary category use, including religiosity which had a significant impact in Study 1. The results of the multilevel multinomial model again indicate that facial features were the most important factor in gender categorization. When face was unambiguously masculine or feminine, participants categorized the target as, respectively, male or female in almost 100% of cases. However, when face was ambiguous, participants' judgments were affected by target's expression and occupation more strongly. Masculine expression paired with ambiguous face led to more male categorization, and vice versa for feminine expression. For occupation, there was a counter-stereotypical effect, with masculine occupation leading to more female categorization and vice versa. In general, participants categorized targets with ambiguous faces as nonbinary more often compared to targets with clearly feminine or masculine faces. Again, an exploratory MCA showed that targets were primarily differentiated by face and resulting gender categorization, contrasting feminine/female vs. masculine/male faces and categorization outcomes, as well as ambiguous/nonbinary vs. unambiguous/binary faces and categorization.

As opposed to Study 1, there was no clear male bias in Study 2. A total of 42% of targets were categorized as male, and 45% as female, so both traditional gender categories were used similarly often. There was even a tendency to categorize ambiguous faces as female (35%, n = 212, of ambiguous faces) more often than male (27%, n = 160, of ambiguous faces). Regarding stereotypes and warmth towards targets, targets categorized as nonbinary were stereotyped as warmer than those categorized as female or male. Categorization had no effect on competence ratings or warmth towards targets.

In Study 2, we addressed some of the limitations from Study 1. The sample was more representative, but still predominantly young (M = 29.0 years, SD = 8.4) and highly educated. Contrary to Study 1, male participants were overrepresented in the sample (63.5% male,

34.5% female, 2.0% diverse), allowing for a more balanced representation across both studies. In Study 2, there were some gender-diverse participants in the sample, but still too few to treat as a separate group for analyses. Our results were robust regardless of whether nonbinary/diverse participants were included in the analysis or not. Percentage of participants with migration background was more representative than in Study 1.

Pretesting the stimulus material allowed us to ensure that stimuli would be perceived as masculine, feminine, or ambiguous as intended. We also matched faces by attractiveness to control attractiveness as a confounding variable. Limitations to Study 2 include the same constraints imposed by self-report measures as in Study 1. We also used only one item to measure warmth and competence each, but since we used the items that most explicitly referred to the measured constructs ("friendly" and "competent", respectively), there is no reason to suspect they did not work as intended. Stimulus material still lacked race diversity since we only used White faces again. Our findings might therefore not apply to different ethnic groups.

General Discussion

With the present studies, we examined how face, occupation, and expression influence gender categorization, including ambiguous presentations of gender. We also examined how the importance of a given feature for categorization shifts depending on the other features presented and whether they are ambiguous or not. Specifically, we focused on how strongly occupation and expression wouuld influence categorization in cases where face is clearly feminine or masculine vs. ambiguous. We also wanted to examine how categorization then impacts warmth and competence stereotypes, and how perceiver characteristics influence categorization. For the latter purpose, we considered gender, political orientation, religiousness, gender identification, ambivalent and modern sexism, gender essentialism, and openness towards nonbinary gender.

Factors in Gender Categorization

Participants used all three available gender categories. Women used the nonbinary category more often although this gender difference was no longer significant when including gender role and attitude variables. This suggests that gender differences in nonbinary categorization may be partially due to differences in gender roles and attitudes such as gender essentialism. Factors such as gender essentialism and traditionalism are also related to anti-trans prejudice according to previous research (Brassel & Anderson, 2020; Broussard & Warner, 2019). In Study 1, less religious participants used the nonbinary category more often as well. No other factor influenced usage of the nonbinary category.

These findings are preliminary, and it remains unclear if/which individual differences influence how often people categorize others as nonbinary. It is noteworthy, however, that participants consistently used the nonbinary category. Providing or omitting semantic information such as category labels shapes participant perception and responses (Tskhay & Rule, 2015). Changing gender categorization paradigms to include other categories in addition or as alternatives to female and male might lead to interesting results and new insights into gender categorization.

Face had the strongest effect on gender categorization in the expected direction, supporting previous findings on its importance in gender categorization (Brown & Perrett, 1993; Deffenbacher et al., 1998; Mangini & Biederman, 2004; Rule & Sutherland, 2017; Zhang et al., 2016), whereas expression and occupation had comparably small and inconsistent effects. When face was unambiguous, gender expression barely influenced categorization. But when face was ambiguous, people considered other cues as well. This resonates with previous findings showing that when primary information, e.g. face, is ambiguous, participants consider secondary information or context cues more strongly in their decision (Huart et al., 2005; Macrae & Martin, 2007; Pernet & Belin, 2012). We extend these findings beyond immediately related information, like hair in addition to face, to more general

information about a person. Still, ambiguous faces have considerable impact on the categorization outcome as well, insofar as that people with ambiguous faces are more likely to be categorized as nonbinary.

Gender expression mainly influenced categorization in the expected direction based on common gender stereotypes. However, occupation had counter-stereotypical effects: masculine occupation often led to more female categorization and vice versa. This finding may be driven by the fact that the sample appeared to be egalitarian in their beliefs and thus perhaps rely less on stereotyping: Participants were mostly nonreligious (Cowling et al., 2019; Herek, 2000; Nagoshi et al., 2008), left-wing (Cowling et al., 2019; Herek, 2000), and young (Herek, 2000). All of these demographic factors are linked to chronic egalitarian beliefs that may have inhibited stereotyping (Johns et al., 2008; Moskowitz et al., 2000; Moskowitz & Li, 2011).

Predicting Evaluations of Warmth and Competence and Warmth towards Targets

Contrary to our hypotheses, gender categorization had no influence on stereotypes or warmth towards targets in most cases. Only in Study 2 were targets categorized as female or male vs. nonbinary rated as less warm. We would have expected that female categorization would lead to highest warmth ratings, followed by nonbinary and then male categorization. No other outcomes were influenced by categorization. Stern and Rule (2018) propose that politically liberal individuals may show no bias against androgynous targets that are difficult to categorize, and in some cases even evaluate them more positively because of the cognitive challenge. This might have applied to our samples which were relatively left-wing on average, and a more positive evaluation of ambiguous targets may have influenced stereotypes and warmth towards targets. Indeed, another hypothesis in our preregistration was that conservative participants would report less warmth toward targets that they found difficult to categorize. We found that categorization difficulty was generally associated with higher reported warmth, but there was a negative interaction between categorization difficulty and

right-wing political views on warmth towards targets. Warmth towards targets that were difficult to categorize was lower in participants with right-wing political views. Participants' political views or underlying need for stability and cognitive closure should be considered as influencing factors in categorization and stereotyping.

We did not further examine gender roles, essentialism, sexism, and attitudes towards nonbinary gender as potential moderators in categorization and stereotyping, since it was outside the scope of the current study. It is plausible that although we did not find a robust main effect of gender categorization on stereotypes and warmth towards targets, that this effect is moderated by individual dispositions such as gender essentialism. Such analyses can be conducted in the future with the datasets we provide here, which are openly available on OSF.

Limitations and Future Directions

Generalizability of the findings is limited since both samples were on average rather young and politically leftist. The convenience sample for Study 1 was less representative than the Prolific sample from Study 2. Usage of exclusively White faces as stimulus material limits generalizability as well, as our results only apply to the perception of White targets. We relied solely on self-report measures that are vulnerable to social desirability, which is, however, less of a problem in factorial survey experiments (Armacost et al., 1991). We only used explicit measures, so we were unable to control for implicit stereotypes and attitudes. We also used single-item measures for warmth towards targets, as well as warmth and competence in Study 2, to prevent participant fatigue from answering multiple items after each of the 9 vignettes. When measuring a "doubly-concrete" construct (i.e., a construct that has a single clear meaning and refers to a clear object) such as warmth towards a person, use of single-item measures is justified (Ang & Eisend, 2018; Bergkvist, 2015). Unfortunately, we were unable to measure reliability of the single-item measures, and sensitivity may also be impacted. However, means and standard deviations of warmth and competence were similar

across both studies, regardless of whether we used a single- or two-item measure for each construct. This implies that the single-item measures still detected meaningful variation as well as the two-item measures.

In future studies, researchers should try to replicate the interaction effects we explored in this study, for example, with a similar factorial survey design optimized for interaction analysis. While we know now that when face gender is ambiguous, additional information like expression influences categorization more strongly, our design was not optimized to uncover the interplay between factors in detail. For example, in a study focusing specifically on categorization of ambiguous faces, the relative importance of different pieces of additional information (e.g., occupation vs. expression) could be examined more closely. To test if the effects we found hold in more realistic scenarios than a factorial survey, virtual reality experiments may be useful, because they allow to freely customize faces of avatars that participants then interact with. They would also allow to present occupation and expression information more implicitly by showing avatars in their workplace or pursuing their hobbies, for example.

Though we still need further research to understand how ambiguous and nonbinary individuals fit into existing models of categorization and stereotyping, we have some preliminary evidence to suggest that they do not simply fall in between feminine/female and masculine/male stereotypes. A follow-up study could include measures of specific ambiguity-related constructs like stereotype incongruency or categorization difficulty (Stern & Rule, 2018) and test if they contribute to nonbinary categorization. More research on the content of gender-ambiguous and/or nonbinary stereotypes could also complement our process-focused work. For example, future studies could gather stereotype content from participants as well as previous studies (e.g. Gazzola & Morrison, 2014; Howansky et al., 2019; K. C. Johnson et al., 2020; Madson, 2000) and theoretical considerations, then compare stereotypes about different gender (e.g., nonbinary, transgender (male/female), cisgender (female/male)) and gender

expression groups (e.g., ambiguous, feminine gender-(non)conforming, masculine gender-(non)conforming) to uncover commonalities and differences.

Practice Implications

Our findings help explain mechanisms behind prejudice and discrimination against TGNC individuals. Negative reactions to TGNC individuals are often related to their gender ambiguity, which can manifest as physical androgyny (Stern & Rule, 2018) or as a perceived mismatch between physical features and self-identification and/or gender expression (Gerhardstein & Anderson, 2010; Morgenroth & Ryan, 2021). Face emerged as the primary factor in gender categorization, and unambiguously masculine or feminine face leads people to categorize the target as male or female, respectively. However, when people apply the same heuristic to trans people who do not "pass" well as their gender and thus categorize them according to their sex assigned at birth, they receive conflicting information through targets' self-identification and/or gender expression. This complicates gender categorization and the ensuing application of stereotypes and social expectations.Similarly, TGNC people with more androgynous faces complicate categorization since the primary basis for categorization is ambiguous, so people have to consider additional information about gender expression and behavior. However, this information is not necessarily diagnostic for gender either, and more subject to contextual variation than face. Therefore, in both cases, perceivers may feel threatened in their worldview and system of thought, as well as in their status, safety, and distinctiveness, depending on what groups they themselves belong to (Morgenroth & Ryan, 2021). Previous studies have shown that, for example, individuals holding more conservative views (Stern & Rule, 2018), traditional gender attitudes (Brassel & Anderson, 2020), and gender-essentialist beliefs (Gallagher & Bodenhausen, 2021) evaluate trans people more negatively, perhaps due to perceived threat to their worldview. The opposite may occur as well – Stern & Rule (2018) hypothesized that people with high ambiguity tolerance and need for cognition may evaluate ambiguous targets more positively.

It should be noted, that actual physical features of the target, interacting with perceiver attitudes, are not the only basis for prejudice and discrimination against TGNC people. Even just labelling a target as trans can negatively influence people's perceptions, again depending on their attitudes as well (Mao et al., 2019). Similarly, a transgender label alone can increase the likelihood of misgendering someone (Howansky et al., 2022), which constitutes a minority stressor for trans individuals and can thus have negative consequences such as feelings of stigma (McLemore, 2015). However, ambiguity of or perceived misalignment between gender-relevant cues could be another contributing factor to this issue.

Conclusion

In two factorial survey experiments, we investigated how face, occupation, and expression and participant characteristics influence gender categorization, and how gender categorization affects evaluations of targets. Face had the largest impact on categorization, especially when it was clearly feminine or masculine. However, when face was ambiguous, occupation and expression influenced participants' categorization decisions more strongly, which highlights that ambiguity of primary information changes how strongly additional information influences categorization. Gender categorization did not predict perceptions of warmth and competence as hypothesized. However, our findings still shed light on gender categorization processes and how they could elicit negative reactions in some individuals in the case of ambiguous gender cues. We showed that overall, ambiguous faces increased likelihood of nonbinary categorization, as well as the influence of other cues besides face in the categorization process, while general readiness to categorize targets as nonbinary varied between individuals.

References

Abele, A. E., Hauke, N., Peters, K., Louvet, E., Szymkow, A., & Duan, Y. (2016). Facets of the fundamental content dimensions: Agency with competence and assertiveness—Communion with warmth and morality. *Frontiers in Psychology*, *7*, 1810. https://doi.org/10.3389/fpsyg.2016.01810

Abele, A. E., & Wojciszke, B. (2014). Communal and agentic content in social cognition: A dual perspective model. In *Advances in Experimental Social Psychology* (Vol. 50, pp. 195–255). Elsevier. https://doi.org/10.1016/B978-0-12-800284-1.00004-7

Adams, K. A., Nagoshi, C. T., Filip-Crawford, G., Terrell, H. K., & Nagoshi, J. L. (2016). Components of gender-nonconformity prejudice. *International Journal of Transgenderism*, *17*(3–4), 185–198. https://doi.org/10.1080/15532739.2016.1200509

American Psychological Association. (2015). Guidelines for psychological practice with transgender and gender nonconforming people. *American Psychologist*, *70*(9), 832–864. https://doi.org/10.1037/a0039906

Ang, L., & Eisend, M. (2018). Single versus multiple measurement of attitudes: A meta-analysis of advertising studies validates the single-item measure approach. *Journal of Advertising Research*, *58*(2), 218–227. https://doi.org/10.2501/JAR-2017-001

Armacost, R. L., Hosseini, J. C., Morris, S. A., & Rehbein, K. A. (1991). An empirical comparison of direct questioning, scenario, and randomized response methods for obtaining sensitive business information. *Decision Sciences*, *22*(5), 1073–1090. https://doi.org/10.1111/j.1540-5915.1991.tb01907.x

Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science*, *7*(3), 136–141. https://doi.org/10.1111/j.1467-9280.1996.tb00346.x

Becker, J. C., & Wagner, U. (2009). Doing gender differently—The interplay of strength of gender identification and content of gender identity in predicting women's endorsement of sexist beliefs. *European Journal of Social Psychology*, *39*(4), 487–508.

https://doi.org/10.1002/ejsp.551

Bennett, D. A. (2001). How can I deal with missing data in my study? *Australian and New Zealand Journal of Public Health*, *25*(5), 464–469. https://doi.org/10.1111/j.1467-842X.2001.tb00294.x

Bergkvist, L. (2015). Appropriate use of single-item measures is here to stay. *Marketing Letters*, *26*(3), 245–255. https://doi.org/10.1007/s11002-014-9325-y

Bodenhausen, G. V, & Peery, D. (2009). Social categorization and stereotyping in vivo: The VUCA challenge. *Social and Personality Psychology Compass*, *3*(2), 133–151. https://doi.org/10.1111/j.1751-9004.2009.00167.x

Brassel, S. T., & Anderson, V. N. (2020). Who Thinks Outside the Gender Box? Feminism, Gender Self-Esteem, and Attitudes toward Trans People. *Sex Roles*, *82*(7–8), 447–462. https://doi.org/10.1007/s11199-019-01066-4

Broussard, K. A., & Warner, R. H. (2019). Gender Nonconformity Is Perceived Differently for Cisgender and Transgender Targets. *Sex Roles*, *80*(7–8), 409–428. https://doi.org/10.1007/s11199-018-0947-z

Broussard, K. A., Warner, R. H., & Pope, A. R. D. (2018). Too Many Boxes, or Not Enough? Preferences for How We Ask About Gender in Cisgender, LGB, and Gender-Diverse Samples. *Sex Roles*, *78*(9–10), 606–624. https://doi.org/10.1007/s11199-017-0823-2

Brown, E., & Perrett, D. I. (1993). What gives a face its gender? *Perception*, *22*(7), 829–840. https://doi.org/10.1068/p220829

Campanella, S., Chrysochoos, A., & Bruyer, R. (2001). Categorical perception of facial gender information: Behavioural evidence and the face-space metaphor. *Visual Cognition*, *8*(2), 237–262. https://doi.org/10.1080/13506280042000072

Canessa-Pollard, V., Reby, D., Banerjee, R., Oakhill, J., & Garnham, A. (2022). The development of explicit occupational gender stereotypes in children: Comparing perceived gender ratios and competence beliefs. *Journal of Vocational Behavior*,

134(February), 103703. https://doi.org/10.1016/j.jvb.2022.103703

Cowling, M. M., Anderson, J. R., & Ferguson, R. (2019). Prejudice-relevant correlates of attitudes towards refugees: A meta-analysis. *Journal of Refugee Studies*, *32*(3), 502–524. https://doi.org/10.1093/jrs/fey062

Deffenbacher, K. A., Hendrickson, C., O'Toole, A. J., Huff, D. P., & Abdi, H. (1998). Manipulating face gender: Effects on categorization and recognition judgments. *Journal of Biological Systems*, *6*(3), 219–239. https://doi.org/10.1142/S0218339098000169

Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2020). Gender Stereotypes Have Changed: A Cross-Temporal Meta-Analysis of U.S. Public Opinion Polls From 1946 to 2018. *American Psychologist*, *75*(3), 301–315. https://doi.org/10.1037/amp0000494.supp

Eckes, T., & Six-Materna, I. (1998). Leugnung von Diskriminierung: Eine Skala zur Erfassung des modernen Sexismus. *Zeitschrift Für Sozialpsychologie*.

Eckes, T., & Six-Materna, I. (1999). Hostilität und Benevolenz: Eine Skala zur Erfassung des ambivalenten Sexismus. *Zeitschrift Für Sozialpsychologie*, *30*, 211–228. https://doi.org/10.1024//0044-3514.30.4.211

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. https://doi.org/10.1037/0022-3514.82.6.878

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science*, *18*(3), 233–239. https://doi.org/10.1111/j.1467-9280.2007.01882.x

Gallagher, N. M., & Bodenhausen, G. V. (2021). Gender essentialism and the mental representation of transgender women and men: A multimethod investigation of stereotype content. *Cognition*, *217*, 1–73.

https://doi.org/10.1016/j.cognition.2021.104887

Ganzeboom, H. B. G. (2010). A new International Socio-Economic Index (ISEI) of occupational status for the International Standard Classification of Occupation 2008 (ISCO-08) constructed with data from the ISSP 2002-2007. *Annual Conference of International Social Survey Programme, Lisbon*, *1*.

Ganzeboom, H. B. G., & Treiman, D. J. (2003). Three internationally standardised measures for comparative research on occupational status. In J. H. P. Hoffmeyer-Zlotnik & C. Wolf (Eds.), *Advances in cross-national comparison* (pp. 159–193). Springer.

Gazzola, S. B., & Morrison, M. A. (2014). Cultural and personally endorsed stereotypes of transgender men and transgender women: Notable correspondence or disjunction? *International Journal of Transgenderism*, *15*(2), 76–99. https://doi.org/10.1080/15532739.2014.937041

Gerhardstein, K. R., & Anderson, V. N. (2010). There's more than meets the eye: Facial appearance and evaluations of transsexual people. *Sex Roles*, *62*(5–6), 361–373. https://doi.org/10.1007/s11199-010-9746-x

Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, *70*(3), 491. https://doi.org/10.1037/0022-3514.70.3.491

Glick, P., Wilk, K., & Perreault, M. (1995). Images of occupations: Components of gender and status in occupational stereotypes. *Sex Roles*, *32*(9–10), 565–582. https://doi.org/10.1007/BF01544212

Goodman, M., Adams, N., Corneil, T., Kreukels, B., Motmans, J., & Coleman, E. (2019). Size and Distribution of Transgender and Gender Nonconforming Populations - A Narrative Review. *Endocrinol Metab Clin N Am*, *48*, 303–321.

Grant, J. M., Mottet, L. A., Tanis, J. T., Harrison, J., Herman, J. L., & Keisling, M. (2011). *Injustice at Every Turn*. National Center for Transgender Equality and National Gay and

Lesbian Task Force. https://dataspace.princeton.edu/jspui/bitstream/88435/dsp014j03d232p/1/NTDS_Report.pdf

Grömping, U. (2018). R Package DoE.base for Factorial Experiments. *Journal of Statistical Software*, *85*(5), 1–41. https://doi.org/10.18637/jss.v085.i05

Haas, A. P., Eliason, M., Mays, V. M., Mathy, R. M., Cochran, S. D., D'Augelli, A. R., Silverman, M. M., Fisher, P. W., Hughes, T., Rosario, M., & others. (2011). Suicide and suicide risk in lesbian, gay, bisexual, and transgender populations: Review and recommendations. *Journal of Homosexuality*, *58*(1), 10–51. https://doi.org/10.1080/00918369.2011.534038

Hamilton, M. C. (1991). Masculine bias in the attribution of personhood: People= male, male= people. *Psychology of Women Quarterly*, *15*(3), 393–402. https://doi.org/10.1111/j.1471-6402.1991.tb00415.x

Hayfield, N. (2013). Never judge a book by its cover?: Students' understandings of lesbian, gay and bisexual appearance. *Psychology \& Sexuality*, *4*(1), 16–24. https://doi.org/10.1080/19419899.2013.748261

Hentschel, T., Heilman, M. E., & Peus, C. V. (2019). The multiple dimensions of gender stereotypes: a current look at men's and women's characterizations of others and themselves. *Frontiers in Psychology*, *10*, 11. https://doi.org/10.3389/fpsyg.2019.00011

Herek, G. M. (2000). The psychology of sexual prejudice. *Current Directions in Psychological Science*, *9*(1), 19–22. https://doi.org/10.1111\%2F1467-8721.00051

Hoenes, J., Sauer, A., & Fütty, T. J. (2019). *Dritte Option beim Geschlechtseintrag für alle?* Bundesvereinigung Trans* e.V. (BVT*). https://www.bundesverband-trans.de/wp-content/uploads/2019/01/dritteOption_V5.pdf

Horn, S. S. (2007). Adolescents' acceptance of same-sex peers based on sexual orientation and gender expression. *Journal of Youth and Adolescence*, *36*(3), 363–371.

https://doi.org/10.1007/s10964-006-9111-0

Howansky, K., Wilton, L. S., Young, D. M., Abrams, S., & Clapham, R. (2019). (Trans) gender stereotypes and the self: Content and consequences of gender identity stereotypes. *Self and Identity*, 1–18. https://doi.org/10.1080/15298868.2019.1617191

Howansky, K., Wittlin, N., Bonagura, D., & Cole, S. (2022). Him, her, them, or none: misgendering and degendering of transgender individuals. *Psychology and Sexuality*, *13*(4), 1026–1040. https://doi.org/10.1080/19419899.2021.1978529

Huart, J., Corneille, O., & Becquart, E. (2005). Face-based categorization, context-based categorization, and distortions in the recollection of gender ambiguous faces. *Journal of Experimental Social Psychology*, *41*(6), 598–608. https://doi.org/10.1016/j.jesp.2004.10.007

International Labour Organization. (2012). *International Standard Classification of Occupations*. International Labour Organization. https://www.ilo.org/public/english/bureau/stat/isco/docs/publication08.pdf

Johar, G. V., Moreau, P., & Schwarz, N. (2003). Gender Typed Advertisements and Impression Formation: The Role of Chronic and Temporary Accessibility. *Journal of Consumer Psychology*, *13*(3), 220–229. https://doi.org/10.1207/S15327663JCP1303_04

Johns, M., Cullum, J., Smith, T., & Freng, S. (2008). Internal motivation to respond without prejudice and automatic egalitarian goal activation. *Journal of Experimental Social Psychology*, *44*(6), 1514–1519. https://doi.org/10.1016/j.jesp.2008.07.003

Johnson, K. C., LeBlanc, A. J., Deardorff, J., & Bockting, W. O. (2020). Invalidation experiences among non-binary adolescents. *The Journal of Sex Research*, *57*(2), 222–233. https://doi.org/10.1080/00224499.2019.1608422

Johnson, K. L., Iida, M., & Tassinary, L. G. (2012). Person (mis) perception: Functionally biased sex categorization of bodies. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1749), 4982–4989. https://doi.org/10.1098/rspb.2012.2060

Karylowski, J. J., Motes, M. A., Wallace, H. M., Harckom, H. A., Hewlett, E. M., Maclean, S. L., & Vaswani, C. L. (2001). Spontaneous gender-stereotypical categorization of trait labels and job labels. *Current Research in Social Psychology*, *6*(6), 77–90.

Koenig, A. M. (2018). Comparing prescriptive and descriptive gender stereotypes about children, adults, and the elderly. *Frontiers in Psychology*, *9*, 1086. https://doi.org/10.3389/fpsyg.2018.01086

Kramer, R. S. S., Jenkins, R., & Burton, A. M. (2017). InterFace: A software package for face image warping, averaging, and principal components analysis. *Behavior Research Methods*, *49*(6), 2002–2011. https://doi.org/10.3758/s13428-016-0837-7

Levy, G. D., Sadovsky, A. L., & Troseth, G. L. (2000). Aspects of young children's perceptions of gender-typed occupations. *Sex Roles*, *42*(11–12), 993–1006. https://doi.org/10.1023/A:1007084516910

Lippa, R. A. (2005). How do lay people weight information about instrumentality, expressiveness, and gender-typed hobbies when judging masculinity-femininity in themselves, best friends, and strangers? *Sex Roles*, *53*(1–2), 43–55. https://doi.org/10.1007/s11199-005-4277-6

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*(3), 86–92. https://doi.org/10.1027/1614-2241.1.3.86

Macrae, C. N., & Martin, D. (2007). A boy primed Sue: Feature-based processing and person construal. *European Journal of Social Psychology*, *37*(5), 793–805. https://doi.org/10.1002/ejsp.406

Madson, L. (2000). Inferences regarding the personality traits and sexual orientation of physically androgynous people. *Psychology of Women Quarterly*, *24*(2), 148–160.

https://doi.org/10.1111/j.1471-6402.2000.tb00196.x

Mangini, M. C., & Biederman, I. (2004). Making the ineffable explicit: Estimating the information employed for face classifications. *Cognitive Science*, *28*(2), 209–226. https://doi.org/10.1207/s15516709cog2802_4

Mao, J. M., Haupert, M. L., & Smith, E. R. (2019). How Gender Identity and Transgender Status Affect Perceptions of Attractiveness. *Social Psychological and Personality Science*, *10*(6), 811–822. https://doi.org/10.1177/1948550618783716

Martin, A. E. (2022). Gender Relativism: How Context Shapes What Is Seen as Male and Female. *Journal of Experimental Psychology: General*, *152*(2), 322–345. https://doi.org/10.1037/xge0001264

Martin, D., & Macrae, C. N. (2007). A face with a cue: Exploring the inevitability of person categorization. *European Journal of Social Psychology*, *37*(5), 806–816. https://doi.org/10.1002/ejsp.445

McLemore, K. A. (2015). Experiences with Misgendering: Identity Misclassification of Transgender Spectrum Individuals. *Self and Identity*, *14*(1), 51–74. https://doi.org/10.1080/15298868.2014.950691

Molin, A., Simond, A., Sato, S., Jaeggi, T., Gygax, P. M., & Meuwly, N. (2020). Linking political and feminist ideology with openness towards non-binary gender: The development and initial validation of a scale to measure subjective Openness towards Non-Binary Gender (ONBG). *Journal of Gender Studies*. https://doi.org/10.1080/09589236.2020.1844644

Morgenroth, T., & Ryan, M. K. (2021). The Effects of Gender Trouble: An Integrative Theoretical Framework of the Perpetuation and Disruption of the Gender/Sex Binary. *Perspectives on Psychological Science*, *16*(6), 1113–1142. https://doi.org/10.1177/1745691620902442

Moskowitz, G. B., & Li, P. (2011). Egalitarian goals trigger stereotype inhibition: A proactive

form of stereotype control. *Journal of Experimental Social Psychology*, *47*(1), 103–116. https://doi.org/10.1016/j.jesp.2010.08.014

Moskowitz, G. B., Salomon, A. R., & Taylor, C. M. (2000). Preconsciously controlling stereotyping: Implicitly activated egalitarian goals prevent the activation of stereotypes. *Social Cognition*, *18*(2), 151–177. https://doi.org/10.1521/soco.2000.18.2.151

Muthén, L. K., & Muthén, B. O. (2010). *Mplus User's Guide* (6th ed.). Muthén & Muthén.

Nadal, K. L., Whitman, C. N., Davis, L. S., Erazo, T., & Davidoff, K. C. (2016). Microaggressions toward lesbian, gay, bisexual, transgender, queer, and genderqueer people: A review of the literature. *The Journal of Sex Research*, *53*(4–5), 488–508. https://doi.org/10.1080/00224499.2016.1142495

Nagoshi, J. L., Adams, K. A., Terrell, H. K., Hill, E. D., Brzuzy, S., & Nagoshi, C. T. (2008). Gender differences in correlates of homophobia and transphobia. *Sex Roles*, *59*(7–8), 521–531. https://doi.org/10.1007/s11199-008-9458-7

Norton, A. T., & Herek, G. M. (2013). Heterosexuals' Attitudes Toward Transgender People: Findings from a National Probability Sample of U.S. Adults. *Sex Roles*, *68*(11–12), 738–753. https://doi.org/10.1007/s11199-011-0110-6

O'Neil, J. M. (2008). Summarizing 25 years of research on men's gender role conflict using the Gender Role Conflict Scale: New research paradigms and clinical implications. *The Counseling Psychologist*, *36*(3), 358–445. https://doi.org/10.1177/0011000008317057

O'Toole, A. J., Deffenbacher, K. A., Valentin, D., McKee, K., Huff, D., & Abdi, H. (1998). The perception of face gender: The role of stimulus structure in recognition and classification. *Memory \& Cognition*, *26*(1), 146–160. https://doi.org/10.3758/BF03211378

Oakes, P. J. (1978). The salience of social categories. In J. C. Turner, M. A. Hogg, P. J. Oakes, S. D. Reicher, & M. S. Wetherell (Eds.), *Rediscovering the social group: A self-categorization theory.* Basil Blackwell.

Oakes, P. J., Turner, J. C., & Haslam, S. A. (1991). Perceiving people as group members: The role of fit in the salience of social categorizations. *British Journal of Social Psychology*, *30*(2), 125–144. https://doi.org/10.1111/j.2044-8309.1991.tb00930.x

Oswald, P. A. (2003). Sex-typing and prestige ratings of occupations as indices of occupational stereotypes. *Perceptual and Motor Skills*, *97*(3), 953–959. https://doi.org/10.2466/pms.2003.97.3.953

Pernet, C. R., & Belin, P. (2012). The role of pitch and timbre in voice gender categorization. *Frontiers in Psychology*, *3*, 23. https://doi.org/10.3389/fpsyg.2012.00023

Plaza, M., Boiché, J., Brunel, L., & Ruchaud, F. (2017). Sport = Male… But Not All Sports: Investigating the Gender Stereotypes of Sport Activities at the Explicit and Implicit Levels. *Sex Roles*, *76*(3–4), 202–217. https://doi.org/10.1007/s11199-016-0650-x

Prolific. (2021). *Prolific*. https://www.prolific.co/

Qualtrics. (2020). *Qualtrics*. https://www.qualtrics.com/

Qualtrics. (2021). *Qualtrics*. https://www.qualtrics.com/

R Core Team. (2013). *R: A Language and Environment for Statistical Computing*. http://www.r-project.org/

RStudio Team. (2020). *RStudio: Integrated Development Environment for R*. http://www.rstudio.com/

Rule, N. O., & Sutherland, S. L. (2017). Social categorization from faces: Evidence from obvious and ambiguous groups. *Current Directions in Psychological Science*, *26*(3), 231–236. https://doi.org/10.1177/0963721417697970

SAS Institute Inc. (2015). *SAS® OnDemand for Academics: User's Guide*. SAS Institute Inc.: Cary, NC.

Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical Methods in Medical Research*, *8*(1), 3–15. https://doi.org/10.1177/096228029900800102

Schwitter, N. (2019). *ISCO08ConveRsions: Converts ISCO-08 to Job Prestige Scores, ISCO-*

88 and Job Name. https://cran.r-project.org/web/packages/ISCO08ConveRsions/ISCO08ConveRsions.pdf

Sinclair, S., & Carlsson, R. (2013). What will I be when I grow up? The impact of gender identity threat on adolescents' occupational preferences. *Journal of Adolescence, 36*(3), 465–474. https://doi.org/10.1016/j.adolescence.2013.02.001

Skewes, L., Fine, C., & Haslam, N. (2018). Beyond Mars and Venus: The role of gender essentialism in support for gender inequality and backlash. *PLoS ONE, 13*(7), 1–17. https://doi.org/10.1371/journal.pone.0200921

Skitka, L. J., & Maslach, C. (1990). Gender roles and the categorization of gender-relevant behavior. *Sex Roles, 22*(3–4), 133–150. https://doi.org/10.1007/bf00288187

Smith, J. S., LaFrance, M., Knol, K. H., Tellinghuisen, D. J., & Moes, P. (2015). Surprising smiles and unanticipated frowns: How emotion and status influence gender categorization. *Journal of Nonverbal Behavior*, *39*(2), 115–130. https://doi.org/10.1007/s10919-014-0202-4

Spizzirri, G., Eufrásio, R., Lima, M. C. P., de Carvalho Nunes, H. R., Kreukels, B. P. C., Steensma, T. D., & Abdo, C. H. N. (2021). Proportion of people identified as transgender and non-binary gender in Brazil. *Scientific Reports*, *11*(1), 1–7. https://doi.org/10.1038/s41598-021-81411-4

Statistisches Bundesamt (Destatis). (2019). *Bevölkerung und Erwerbstätigkeit*. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Migration-Integration/Publikationen/Downloads-Migration/migrationshintergrund-2010220187004.pdf?__blob=publicationFile

Stern, C., & Rule, N. O. (2018). Physical androgyny and categorization difficulty shape political conservatives' attitudes toward transgender people. *Social Psychological and Personality Science, 9*(1), 24–31. https://doi.org/10.1177/1948550617703172

Stroessner, S. J., Haines, E. L., Sherman, J. W., & Kantrowitz, C. J. (2010). Stereotype

relevance moderates category activation: Evidence from the indirect category accessibility task (ICAT). *Social Psychological and Personality Science*, *1*(4), 335–343. https://doi.org/10.1177/1948550610372144

SurveyCircle. (2020). *SurveyCircle*. https://www.surveycircle.com

Swim, J. K., & Hyers, L. L. (2009). Sexism. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 407–430). Taylor & Francis.

Taylor, J. (2003). Women's Leisure Activities, Their Social Stereotypes and Some Implications for Identity. *British Journal of Occupational Therapy*, *66*(4), 151–158. https://doi.org/10.1177/030802260306600404

Tierney, N., Cook, D., McBain, M., & Fay, C. (2020). *naniar: Data Structures, Summaries, and Visualisations for Missing Data*. https://cran.r-project.org/package=naniar

Troche, S. J., & Rammsayer, T. H. (2011). Eine Revision des deutschsprachigen Bem Sex-Role Inventory. *Klinische Diagnostik Und Evaluation*, *4*, 262–283.

Tskhay, K. O., & Rule, N. O. (2015). Semantic information influences race categorization from faces. *Personality and Social Psychology Bulletin*, *41*(6), 769–778. https://doi.org/10.1177/0146167215579053

White, M. J., Kruczek, T. A., Brown, M. T., & White, G. B. (1989). Occupational sex stereotypes among college students. *Journal of Vocational Behavior*, *34*(3), 289–298. https://doi.org/10.1016/0001-8791(89)90021-3

White, M. J., & White, G. B. (2006). Implicit and explicit occupational gender stereotypes. *Sex Roles*, *55*(3–4), 259–266. https://doi.org/10.1007/s11199-006-9078-z

Worthen, M. G. F., Lingiardi, V., & Caristo, C. (2017). The Roles of Politics, Feminism, and Religion in Attitudes Toward LGBT Individuals: A Cross-Cultural Study of College Students in the USA, Italy, and Spain. *Sexuality Research and Social Policy*, *14*(3), 241–258. https://doi.org/10.1007/s13178-016-0244-y

Zarate, M. A., & Smith, E. R. (1990). Person categorization and stereotyping. *Social*

Cognition, *8*(2), 161–185. https://doi.org/10.1521/soco.1990.8.2.161

Zhang, X., Li, Q., Sun, S., & Zuo, B. (2016). The time course from gender categorization to gender-stereotype activation. *Social Neuroscience*, *13*(1), 52–60. https://doi.org/10.1080/17470919.2016.1251965

Zhang, X., Li, Q., Sun, S., & Zuo, B. (2019). Facial expressions can inhibit the activation of gender stereotypes. *Cognition and Emotion*, *33*(7), 1424–1435. https://doi.org/10.1080/02699931.2019.1586648

Zhao, M., & Hayward, W. G. (2010). Holistic processing underlies gender judgments of faces. *Attention, Perception, \& Psychophysics*, *72*(3), 591–596. https://doi.org/10.3758/APP.72.3.591

Table 1

Dimensions of gender categorization and levels of each dimension in Study 1 design

Dimension	Levels
Face	(1) masculine, (2) ambiguous, (3) feminine
Expression	(1) never wears make up and does not like dressing up, (2) sometimes wears make up and sometimes likes dressing up, (3) always wears make up and likes dressing up a lot
Occupation	(1) firefighter/athletic trainer/police officer, (2) restaurant manager/travel agent/photographer, (3) beautician/dental hygienist/kindergarten teacher

Table 2

Gender categorization results by participant gender in Study 1

Participant gender	Nonbinary		Female		Male		Total
	n	%	n	%	n	%	n
Female	161	11.4	556	39.3	696	49.3	1413
Male	24	6.2	162	41.9	201	51.9	387
Total	185	10.3	718	39.9	897	49.8	1800

Note. Number of participants = 200, total number of observations = 1800.

Table 3

Predicting the use of the nonbinary category from participant characteristics, Study 1

	Step 1						Step 2					
Effect	Estimate	SE	95% CI		p	OR	Estimate	SE	95% CI		p	OR
			LL	UL					LL	UL		
Count model												

Participant gender (f vs. m)	.72	.29	.14	1.29	.015	2.05	.69	.30	.10	1.29	.022
Age	.01	.01	-.01	.02	.347	1.01	.01	.01	.00	.02	.200
Political views	.02	.07	-.12	.16	.758	1.02	.06	.06	-.06	.19	.311
Religiosity	-.13	.05	-.23	-.04	.005	0.88	-.14	.04	-.22	-.06	<.001
Gender identification							-.04	.11	-.25	.18	.745
Ambivalent sexism							-.21	.15	-.50	.08	.151
Modern sexism							.00	.15	-.30	.30	.993
Sex role (Amb/undiff vs. masc/fem)							-.02	.18	-.38	.33	.894
Zero-inflation model											

Participant gender (m vs. f)	.70	1.35	-1.95	3.35	.604	2.01	1.34	1.24	-1.09	3.76	.281
Age	-.03	.04	-.10	.04	.405	0.97	-.05	.04	-.12	.02	.180
Political views	.47	.25	-.02	.96	.058	1.60	.41	.23	-.04	.87	.073
Religiosity	-.76	.54	-1.83	.30	.161	0.47	-.96	.48	-1.90	-.02	.046
Gender identification							.07	.36	-.63	.77	.846
Ambivalent sexism							.00	.57	-1.11	1.11	.995
Modern sexism							.77	.61	-.43	1.97	.208
Sex role (amb/undiff vs. masc/fem)							1.10	.83	-.53	2.74	.185

Note. Number of participants = 198. OR – odds ratio.

Table 4

Influence of facial features, occupation, and behavior on nonbinary vs. female/male categorization in Study 1

Effect	Estimate	SE	95% CI		p	OR
			LL	UL		
Nonbinary vs. Male						
A vs. M face	2.25	.26	1.74	2.77	<.001	9.49
A vs. M occupation	-0.31	.18	-.65	0.04	.085	0.73
A vs. M behavior	0.77	.21	0.36	1.17	<.001	2.16
Nonbinary vs. Female						
A vs. F face	3.95	.36	3.25	4.65	<.001	51.94
A vs. F occupation	-0.52	.23	-0.97	-0.07	.025	0.59

A vs. F behavior	-0.68	.22	-.11	-.24	.002	0.51

Note. A = ambiguous, M = masculine, F = feminine. We only report matching contrasts (e.g., ambiguous vs. male for non-binary vs. male categorization) for clarity, for full table see Table 5.

Number of participants = 214, total number of observations = 1838.

Table 5

Influence of facial features, occupation, and behavior, including interactions, on male/female vs. nonbinary categorization, Study 1

Effect	Estimate	SE	95% CI		p
			LL	UL	
Male vs. nonbinary					
M vs. A face	3.32	0.61	2.14	4.51	<.001
F vs. A face	-0.28	1.48	-3.18	2.63	.853
M vs. A job	-0.35	0.43	-1.19	0.49	.412
F vs. A job	0.62	0.36	-0.09	1.33	.088
M vs. A behavior	1.46	0.43	0.62	2.29	.001
F vs. A behavior	-0.06	0.36	-0.77	0.65	.866

M face x M job	-0.85	0.56	-1.95	0.25	.131
M face x F job	-1.46	0.53	-2.50	-0.42	.006
M face x M behavior	1.15	1.09	-0.99	3.28	.292
M face x F behavior	-0.87	0.49	-1.83	0.09	.075
F face x M job	8.44	1.57	5.36	11.52	<.001
F face x F job	-0.74	1.11	-2.92	1.44	.505
F face x M behavior	-1.52	1.50	-4.47	1.42	.311
F face x F behavior	-1.18	1.37	-3.87	1.51	.389
M job x M behavior	-0.99	0.59	-2.15	0.17	.093
M job x F behavior	0.91	0.60	-0.26	2.09	.128
F job x M behavior	-0.91	0.62	-2.13	0.31	.142
F job x F behavior	0.60	0.50	-0.39	1.59	.233
Female vs. nonbinary					
M vs. A face	-0.41	0.77	-1.93	1.10	.593
F vs. A face	4.54	1.13	2.33	6.75	<.001
M vs. A job	-0.07	0.43	-0.91	0.77	.870

F vs. A job	-1.48	0.48	-2.42	-0.53	.002
M vs. A behavior	-0.92	0.49	-1.89	0.05	.062
F vs. A behavior	-1.16	0.44	-2.03	-0.30	.009
M face x M job	-1.98	1.07	-4.08	0.12	.065
M face x F job	-0.70	1.03	-2.71	1.31	.495
M face x M behavior	2.51	1.37	-0.17	5.19	.067
M face x F behavior	0.09	0.92	-1.72	1.89	.926
F face x M job	9.01	0.55	7.94	10.08	<.001
F face x F job	-0.33	0.61	-1.52	0.86	.587
F face x M behavior	-0.99	1.13	-3.21	1.23	.383
F face x F behavior	-0.70	1.17	-3.00	1.60	.551
M job x M behavior	0.92	0.65	-0.34	2.19	.154
M job x F behavior	0.40	0.70	-0.97	1.77	.564
F job x M behavior	1.71	0.73	0.28	3.13	.019
F job x F behavior	1.63	0.73	0.19	3.06	.026

Note. M = masculine, F = feminine, A = ambiguous. Number of participants = 214, Number of observations = 1838.

Table 6

Warmth, competence, and attitude towards the target predicted from gender categorization in Study 1

Effect	Estimate	SE	95% CI		p
			LL	UL	
Warmth					
Nonbinary vs. female	-.04	.07	-.18	.09	.535
Nonbinary vs. male	<.01	.07	-.13	.14	.985
Competence					
Nonbinary vs. female	-.02	.07	-.15	.11	.734
Nonbinary vs. male	.05	.06	-.07	.17	.429
Attitude					
Nonbinary vs. female	-.30	.17	-.63	.03	.077
Nonbinary vs. male	.06	.18	-.29	.41	.740

Note. Number of participants = 214, total number of observations = 1838.

Table 7

Face, behavior, and job predicting stereotypes, Study 1

Outcome	Predictor	*b*	*b* 95% CI [LL, UL]	*beta*	*beta* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*r*	Fit
Warmth	(Intercept)	3.33**	[3.21, 3.45]						
	face_fem	0.07	[-0.04, 0.19]	0.04	[-0.02, 0.09]	.00	[-.00, .00]	.03	
	face_m	0.04	[-0.07, 0.15]	0.02	[-0.03, 0.07]	.00	[-.00, .00]	.00	
	job_fem	0.04	[-0.07, 0.15]	0.02	[-0.03, 0.07]	.00	[-.00, .00]	.02	
	job_m	-0.01	[-0.12, 0.10]	-0.01	[-0.06, 0.05]	.00	[-.00, .00]	-.02	
	behav_fem	0.07	[-0.04, 0.18]	0.03	[-0.02, 0.09]	.00	[-.00, .00]	.03	
	behav_m	0.03	[-0.09, 0.14]	0.01	[-0.04, 0.07]	.00	[-.00, .00]	-.00	
									R^2 = .002
									95% CI[.00,.00]

Competence	(Intercept)	3.53**	[3.42, 3.63]					
	face_fem	-0.03	[-0.13, 0.06]	-0.02	[-0.07, 0.03]	.00	[-.00, .00]	.00
	face_m	-0.08	[-0.18, 0.01]	-0.05	[-0.10, 0.01]	.00	[-.00, .01]	-.04
	job_fem	-0.02	[-0.12, 0.08]	-0.01	[-0.06, 0.04]	.00	[-.00, .00]	-.02
	job_m	0.03	[-0.06, 0.13]	0.02	[-0.03, 0.07]	.00	[-.00, .00]	.02
	behav_fem	-0.01	[-0.10, 0.09]	-0.00	[-0.06, 0.05]	.00	[-.00, .00]	.01
	behav_m	-0.05	[-0.15, 0.04]	-0.03	[-0.08, 0.02]	.00	[-.00, .00]	-.03
								R^2 = .003
								95% CI[.00,.01]
Attitude	(Intercept)	4.76**	[4.50, 5.02]					
	face_fem	0.17	[-0.07, 0.41]	0.04	[-0.02, 0.09]	.00	[-.00, .00]	.06**
	face_m	-0.23	[-0.47, 0.01]	-0.05	[-0.10, 0.00]	.00	[-.00, .01]	-.07**
	job_fem	-0.03	[-0.27, 0.21]	-0.01	[-0.06, 0.05]	.00	[-.00, .00]	.00
	job_m	-0.08	[-0.32, 0.16]	-0.02	[-0.07, 0.03]	.00	[-.00, .00]	-.01
	behav_fem	-0.06	[-0.30, 0.18]	-0.01	[-0.07, 0.04]	.00	[-.00, .00]	-.02
	behav_m	0.03	[-0.21, 0.27]	0.01	[-0.05, 0.06]	.00	[-.00, .00]	.01

	R^2 = .007
	95% CI[.00,.01]

Note. A significant *b*-weight indicates the beta-weight and semi-partial correlation are also significant. *b* represents unstandardized regression weights. *beta* indicates the standardized regression weights. sr^2 represents the semi-partial correlation squared. *r* represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.

* indicates $p < .05$. ** indicates $p < .01$.

Table 8

Dimensions of gender categorization and levels of each dimension in Study 2

Dimension	Levels
Face	(1) masculine, (2) ambiguous, (3) feminine
Expression	(1) bodybuilding/rugby/home improvement (2) swimming/running/cooking (3) riding/cheerleading/embroidery
Occupation	(1) firefighter/construction worker/forester (2) photographer/psychologist/musician (3) beautician/kindergarten teacher/nurse

Table 9

Gender categorization results by participant gender in Study 2

Participant	Nonbinary		Female		Male		Total
gender	n	%	n	%	n	%	N
Female	107	17.2	252	40.6	262	42.2	621
Male	126	11.0	533	46.6	484	42.3	1143
Nonbinary	7	19.4	16	44.4	13	36.1	36
Total	240	13.3	801	44.5	759	42.2	1800

Note. Number of participants = 204, total number of observations = 1800.

Table 10

Predicting the use of the nonbinary category from participant characteristics, Study 2

	Step 1					Step 2				
Effect	Estimate	SE	95% CI	p	OR	Estimate	SE	95% CI	p	OR

			LL	UL					LL	UL		
Count model												
Participant gender (f vs. m)	.33	.16	.02	.65	.038	1.39	.29	.17	-.05	.63	.094	1.34
Age	.00	.01	-.01	.02	.759	1.00	.00	.01	-.02	.02	.893	1.00
Political views	.07	.05	-.02	.16	.125	1.07	.09	.06	-.02	.20	.094	1.09
Religiosity	.02	.03	-.04	.09	.480	1.02	.02	.03	-.04	.08	.545	1.02
Gender identification							-.06	.08	-.21	.09	.408	0.94
Openness towards nonbinary gender							.00	.07	-.14	.15	.960	1.00
Modern sexism							-.10	.11	-.31	.10	.330	0.90
Gender essentialism							.02	.09	-.17	.20	.839	1.02
Zero-inflation model												

Participant gender (m vs. f)	-.87	1.04	-2.90	1.17	.405	0.42	-.69	1.47	-3.58	2.20	.641	0.50
Age	.03	.05	-.07	.13	.532	1.03	.00	.06	-.11	.11	.989	1.00
Political views	.61	.39	-.15	1.36	.115	1.84	.23	.52	-.79	1.24	.661	1.26
Religiosity	.13	.14	-.15	.41	.366	1.14	.22	.19	-.15	.58	.251	1.25
Gender identification							.13	.46	-.78	1.03	.779	1.14
Openness towards nonbinary gender							-.14	.42	-.96	.68	.734	0.87
Modern sexism							.94	.56	-.16	2.03	.094	2.56
Gender essentialism							.42	.64	-.84	1.68	.515	1.52

Note. Number of participants = 198. OR – odds ratio.

Table 11

Influence of facial features, occupation, and behavior on nonbinary vs. female/male categorization in Study 2

Effect	Estimate	SE	95% CI		p	Odds ratio
			LL	UL		
Nonbinary vs. Male						
A vs. M face	6.14	0.71	4.75	7.52	<.001	464.05
A vs. M occupation	-0.14	0.23	-0.60	0.31	.540	0.87
A vs. M behavior	0.42	0.20	0.04	0.81	.029	1.52
Nonbinary vs. Female						
A vs. F face	4.34	0.36	3.64	5.04	<.001	76.71
A vs. F occupation	-0.32	0.18	-0.67	0.04	.079	0.73
A vs. F behavior	0.44	0.20	0.04	0.83	.030	1.55

Note. A = ambiguous, M = masculine, F = feminine. We only report matching contrasts (e.g., ambiguous vs. male for non-binary vs. male categorization) for clarity, for full table see Table 12.

Number of participants = 214, total number of observations = 1838.

Table 12

Influence of facial features, occupation, and behavior, including interactions, on male/female vs. nonbinary categorization, Study 2

Effect	Estimate	SE	95% CI		p
			LL	UL	
Male vs. nonbinary					
M vs. A face	4.52	1.06	2.44	6.60	<.001
F vs. A face	-10.32	0.75	-11.79	-8.84	<.001
M vs. A job	-0.89	0.58	-2.02	0.25	.125
F vs. A job	-0.23	0.40	-1.02	0.57	.574
M vs. A behavior	-0.11	0.40	-0.90	0.67	.779
F vs. A behavior	0.79	0.47	-0.12	1.71	.089
M face x M job	12.81	1.23	10.41	15.21	<.001

M face x F job	0.21	1.50	-2.73	3.15	.887
M face x M behavior	12.60	0.80	11.04	14.16	<.001
M face x F behavior	11.48	0.82	9.88	13.08	<.001
F face x M job	-10.51	1.32	-13.10	-7.91	<.001
F face x F job	0.23	2.46	-4.58	5.05	.925
F face x M behavior	9.72	1.44	6.90	12.54	<.001
F face x F behavior	9.42	2.45	4.63	14.22	<.001
M job x M behavior	1.68	0.75	0.20	3.15	.026
M job x F behavior	-1.07	0.86	-2.76	0.62	.213
F job x M behavior	0.08	0.62	-1.14	1.30	.899
F job x F behavior	-0.58	0.68	-1.92	0.75	.391
Female vs. nonbinary					
M vs. A face	-9.09	1.12	-11.27	-6.90	<.001
F vs. A face	5.06	0.84	3.42	6.71	<.001
M vs. A job	2.34	0.52	1.32	3.36	<.001

F vs. A job	0.98	0.48	0.04	1.92	.042
M vs. A behavior	0.82	0.52	-0.20	1.83	.116
F vs. A behavior	2.69	0.53	1.66	3.72	<.001
M face x M job	19.33	1.57	16.26	22.40	<.001
M face x F job	-0.84	1.45	-3.69	2.00	.561
M face x M behavior	11.14	0.90	9.38	12.91	<.001
M face x F behavior	2.40	1.46	-0.45	5.26	.099
F face x M job	0.53	1.24	-1.90	2.97	.668
F face x F job	0.71	1.34	-1.92	3.33	.598
F face x M behavior	-1.00	1.38	-3.71	1.71	.467
F face x F behavior	-1.55	1.67	-4.82	1.73	.354
M job x M behavior	-3.03	0.83	-4.64	-1.41	<.001
M job x F behavior	-3.12	0.73	-4.56	-1.69	<.001
F job x M behavior	-0.89	0.74	-2.33	0.56	.231
F job x F behavior	-2.98	0.75	-4.44	-1.52	<.001

Note. M = masculine, F = feminine, A = ambiguous. Number of participants = 200, Number of observations = 1800.

Table 13

Warmth, competence, and attitude towards the target predicted from gender categorization in Study 2

Effect	Estimate	SE	95% CI		p
			LL	UL	
Warmth					
Nonbinary vs. female	.20	.06	.07	.32	.002
Nonbinary vs. male	.18	.06	.07	.29	.001
Competence					
Nonbinary vs. female	-.08	.06	-.20	.04	.215

Nonbinary vs. male	-.10	.06	-.22	.02	.116
Attitude					
Nonbinary vs. female	-.21	.19	-.58	.15	.249
Nonbinary vs. male	-.01	.17	-.35	.33	.952

Note. Number of participants = 200, total number of observations = 1800.

Figure 1. Example of the stimuli presenting information about face, expression, and occupation.

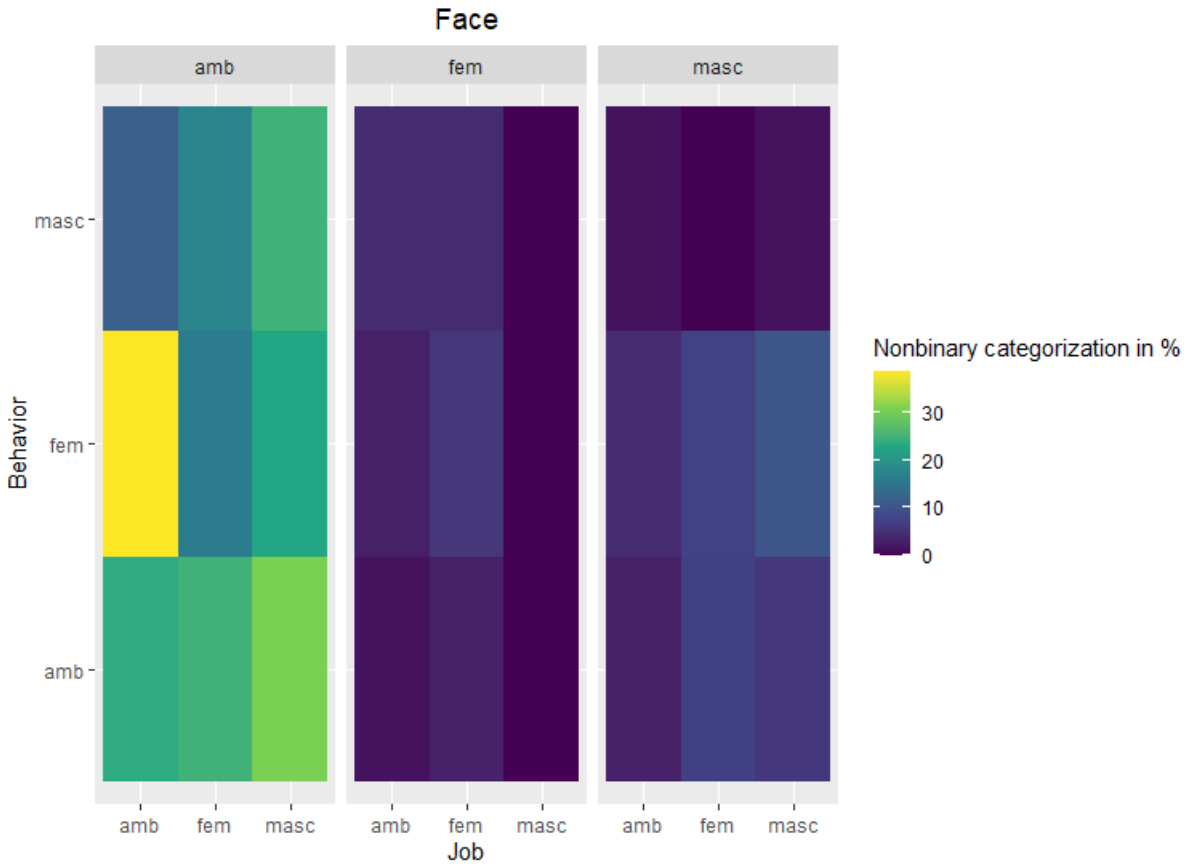
This person is an athletic trainer.

This person always wears makeup and likes to dress up.

What do you think is the gender of this person?

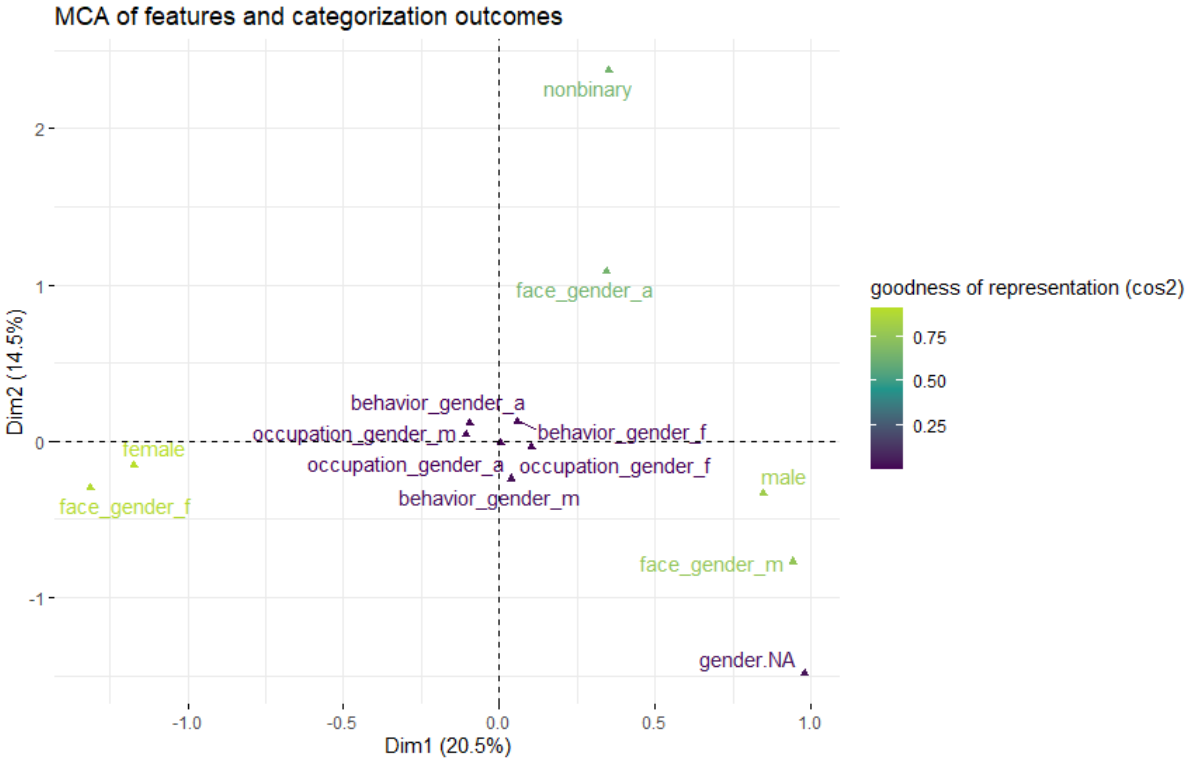
- male
- female
- diverse/nonbinary

Figure 2. *Percentage of targets categorized as nonbinary depending on face, behavior, and job in Study 1*



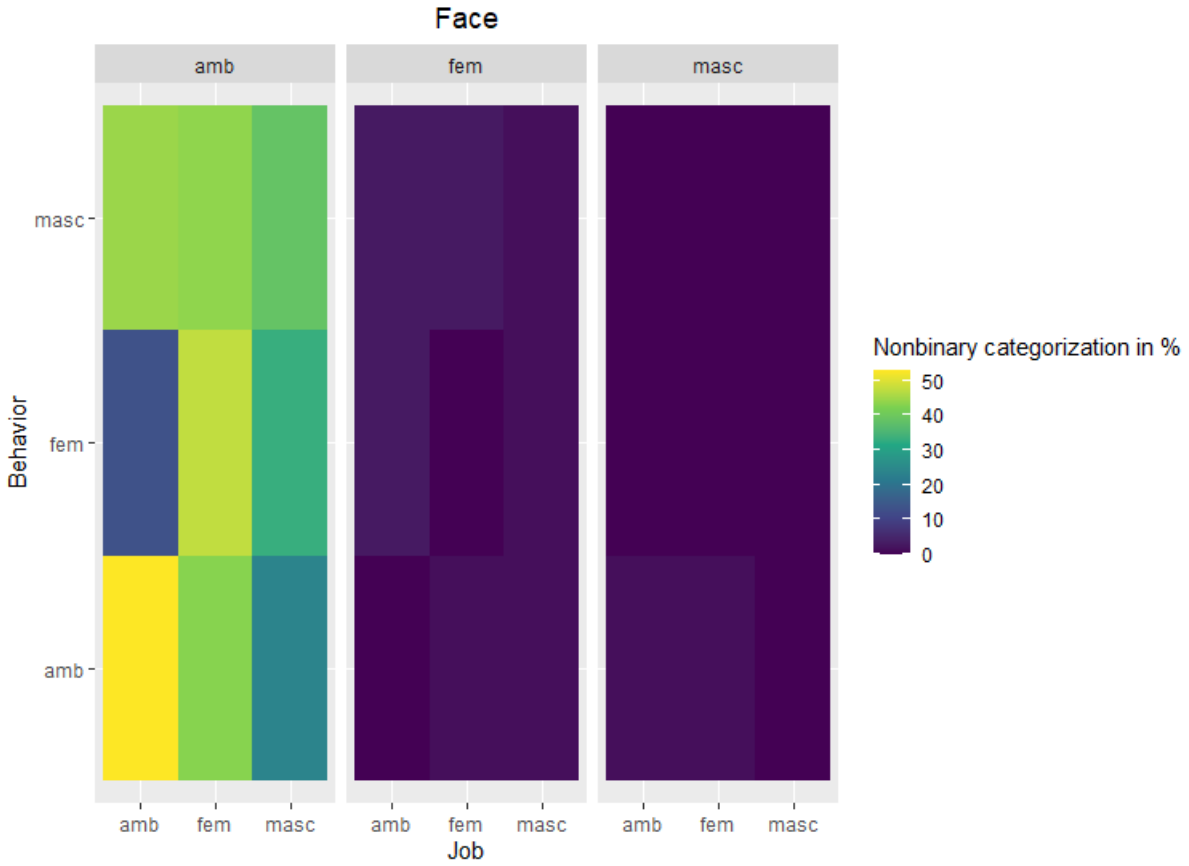
Note. amb = ambiguous, fem = feminine, masc = masculine. For specific percentages, see Online Resource 1, Table 6.

Figure 3. *Multiple correspondence analysis of features and categorization outcomes in Study 1*



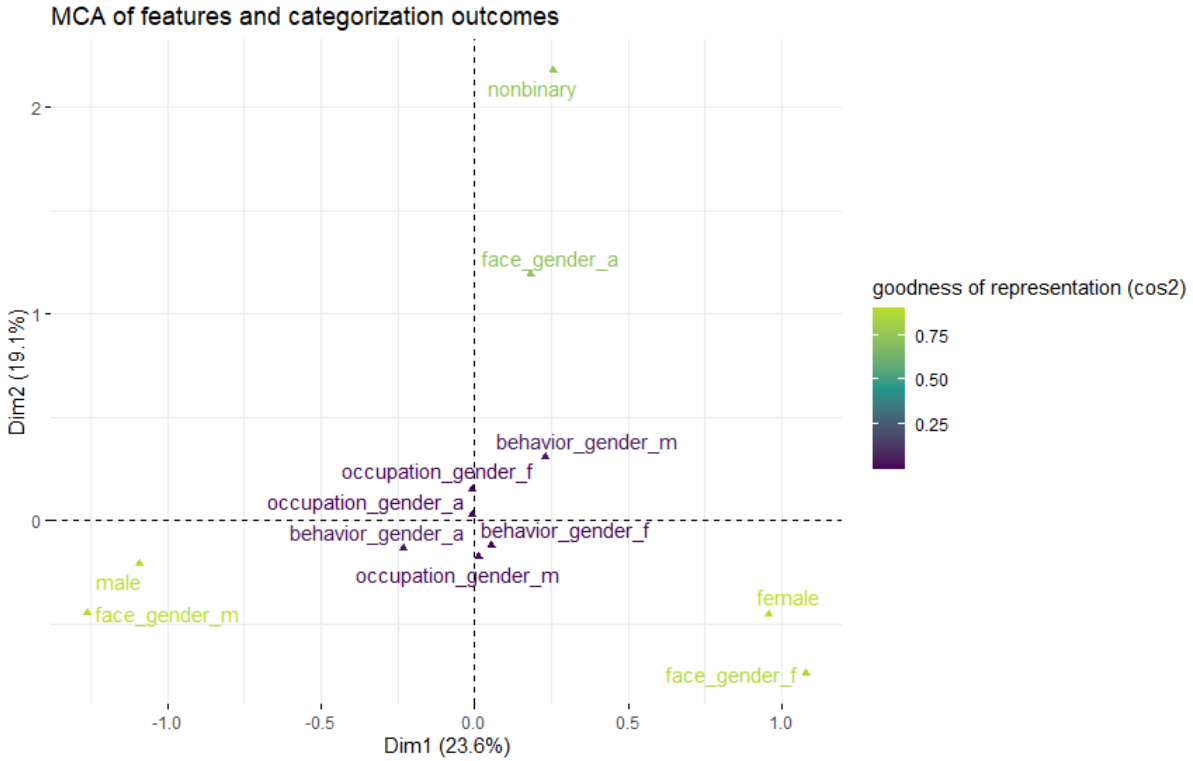
Note. a = ambiguous, f = feminine, m = masculine.

Figure 4. *Percentage of targets categorized as nonbinary depending on face, behavior, and job in Study 2*



Note. amb = ambiguous, fem = feminine, masc = masculine. For specific percentages, see Online Resource 1, Table 7.

Figure 5. *Multiple correspondence analysis of features and categorization outcomes in Study 2*



Note. a = ambiguous, f = feminine, m = masculine.