Journal of Experimental Social Psychology 116 (2025) 104691

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Registered Report Stage II

Gender categorization and memory in transgender and cisgender people☆

Natalie M. Gallagher [a,*], Emily Foster-Hanson [b], Kristina R. Olson [a]

[a] Princeton University, Department of Psychology, Peretsman Scully Hall, Princeton, NJ 08540, United States of America
[b] Swarthmore College, Department of Psychology, 500 College Ave, Swarthmore, PA 19070, United States of America

ARTICLE INFO

Keywords:
Gender
Categorization
Person memory
Transgender
Gender modality

ABSTRACT

Gender categorization is central to everyday life. Discussions about gender have traditionally focused on *gender identities*, or gender categories to which a person might have an internal sense of belonging (e.g., men and women, boys and girls). More recently, discussions about gender also include *gender modality* (transgender or cisgender), or how a person's gender identity relates to their sex assigned at birth. In this registered report, we investigate gender-relevant categorization including gender identity and gender modality using measures assessing the automatic encoding of categories and explicit beliefs about the similarity between categories. We also compare performance on these tasks in transgender and cisgender youth and adults to help shed light on long-standing debates about the role of experience in categorization. Across two studies ($N = 1144$), we found that participants automatically encoded both gender identity and gender modality, and that variations in categorization between participant groups were largely mediated by participants' attitudes (i.e., openness to nonbinary identities) and experiences (i.e., contact with trans people). These results thus help refine our psychological theories of gender categorization to more accurately reflect the landscape of gender categories permeating modern society.

Gender categorization is central to everyday life, shaping everything from the language we use to describe the people around us (as *he*, *she*, or *they*), to the bathroom where we direct a stranger, to which TSA agent pats down which traveler at the airport. Both in broader society and the psychological study of gender categories, discussions about gender have traditionally focused on binary divisions between men and women, boys and girls. These categories refer to different *gender identities*, or gender categories to which a person might have an internal sense of belonging.[1] However, more recently, there's been increasing discussion about another dimension of gender, which we refer to here as *gender modality* (transgender or cisgender; Ashley, 2022), or how a person's gender identity is related to their gender assigned at birth. Gender modality is an increasingly common topic in popular discourse, particularly among young people (Jones, 2021), but very little research on gender categorization to date has focused on this dimension. Current theories about how people categorize the gender of other people are thus underspecified with respect to gender modality.

In the present study, we investigate gender-relevant categorization along dimensions of both gender identity (men and women) and gender modality (transgender and cisgender). We investigate two aspects of categorization: (1) the automatic encoding of gender-relevant categories and (2) explicit beliefs about the similarity between gender categories. Exploring these questions helps to refine our psychological theories of gender categorization so that they more accurately reflect the landscape of gender categories that permeate modern society. By including both implicit and explicit measures and a within-participants design, this work can further illuminate the relation between implicit gender categorization and explicit beliefs about gender at the individual level.

The current work also compares these processes at the group level, by including groups of participants who systematically vary in their experiences with gender: transgender and cisgender youth and adults. These comparisons can help shed light on long-standing debates about the role of experience in categorization more broadly (Bigler & Liben, 2007). For example, does a perceiver's own gender modality shape their gender categorization? Furthermore, age differences between young people and adults in their familiarity/contact with transgender people

☆ This paper has been recommended for acceptance by Evava Pietri.

* Corresponding author.
E-mail address: nmgallag@princeton.edu (N.M. Gallagher).

[1] In addition, an increasing number of people are using labels like nonbinary and genderqueer to describe their gender. In this work we do not include these identities because this task focuses on stereotypic presentations of gender identity, and not enough psychological research on lay representations of these identities has been done to include those representations in the kinds of tasks we use in our research. We return to this as a limitation and area for future work in the discussion.

https://doi.org/10.1016/j.jesp.2024.104691
Received 2 April 2023; Received in revised form 8 October 2024; Accepted 18 October 2024
Available online 30 October 2024

(Jones, 2021; Minkin & Brown, 2021) provide a unique test of how categorization might vary across groups with different experiences. To examine the role of experience, we ask whether gender-relevant categorization varies by participant gender modality (transgender or cisgender) and participant age group (youth or adults). We also test whether group differences are related to differences in familiarity with transgender people and views of gender fluidity. Through this work we hope to shed light on how people process gender-relevant categories in daily life, and the life experiences that influence those processes.

1.1. Gender categorization

Categorization is a fundamental feature of human cognition: knowing what category something belongs to allows us to simplify our world and make the unfamiliar familiar (Mervis & Pani, 1980; Murphy, 2002; Rosch, 1978). When making sense of the social world, gender is the earliest-emerging and most ubiquitous social category that people use (Maccoby, 1988; Martin & Ruble, 2004; M. G. Taylor, 1996). People learn to categorize themselves and those around them by gender beginning in infancy (Cohen & Strauss, 1979; Leinbach & Fagot, 1993; Quinn et al., 2002; Younger & Fearing, 1999). Preschool-aged children label themselves and others as belonging to gender categories (Weinraub et al., 1984; Zosuls et al., 2009) and they use gender to guide their behavior—from how they play (Ruble et al., 2006; Zosuls et al., 2009) to which career interests they pursue (Bian et al., 2017; Kung, 2021). Children even prioritize grouping people by gender over other social categories (e.g., race; Shutts et al., 2013).

By adulthood, gender categorization is so rapid and implicit that people find it hard to avoid categorizing the people they encounter by gender (Fiske, 1998). The most classic demonstration of this effect is the *Who-Said-What?* Task (S. E. Taylor et al., 1978), in which participants observe photographs of individuals who vary by gender identity, paired with statements supposedly said by those individuals. Later, when subjects are asked to recall who said each statement, they often make non-random errors. People systematically misremember statements as being said by others who share the speaker's gender identity, while cross-identity errors are much less common. This task has been interpreted as evidence that people automatically encode gender, because participants were not informed that the task was related to gender yet they categorized along this dimension anyway. Children as young as three years old show the Who-Said-What effect in the domain of gender (Weisman et al., 2015). This effect has been observed for other social categories such as race (Sesko & Biernat, 2010), social class (Weeks & Lupfer, 2004), and even mask-wearing or other indicators of arbitrary or coalitional social groups (Castelli et al., 2022; Kurzban et al., 2001).

1.2. Categorization of transgender targets

Despite the robust evidence that people use gender categories to make sense of the social world throughout daily life, nearly all research on gender categorization to date has focused only on *gender identity* (e.g., whether someone identifies as a man or woman). The boundaries of gender identities have sometimes been explored–for example, showing that people are less accurate in categorizing masculine women and feminine men than they are in categorizing more prototypically feminine women and masculine men into the categories of "men" and "women" (Strauss et al., 2012). However, this work has largely focused on cisgender targets.

Only in the last decade have researchers begun to assess how people categorize or think about categories of people who vary by gender modality. For example, one recent study found that stereotypes of transgender men and women were more overlapping with each other than the stereotypes of either group were with cisgender people who shared their gender identities or sex assigned at birth (Howansky et al., 2019). This suggests that at least in some contexts, categorization by gender modality can even be stronger than by gender identity.

In the last few years, more attention has been paid to how people categorize the faces of transgender people. Conceptually replicating findings from cisgender targets, Stern and Rule (2018), for example, demonstrated that adults are slower to categorize the gender identity of more androgynous transgender people than less androgynous transgender people. Perceivers also judge people with androgynous faces more negatively (Broussard & Warner, 2019; Gerhardstein & Anderson, 2010; Stern & Rule, 2018), especially when they present themselves as more typical of their gender identity (i.e., trans women with long hair) because they're perceived as transgressing binary gender boundaries (Broussard & Warner, 2019).

Studies have also examined perceptions of faces when they are *labelled* as transgender. For example, Mao et al. (2019), observed that faces described as transgender were perceived as less attractive than those same faces when they were described as cisgender. The most consistent finding in this literature is that labeling faces as cisgender or transgender impacts how masculine or feminine a person is perceived to be. In a series of studies, Wittlin et al. (2018) found that when people were told that a person was transgender, they perceived that person to be more androgynous than when the same person was described as cisgender. Howansky et al. (2020) expanded on this finding by showing that participants made more-androgynous avatars to represent faces of people who were labelled as transgender than to represent those same faces labelled as cisgender. They also found that perceiving transgender individuals' faces as less gender typical was associated with greater discomfort with the individuals expressing themselves in gender typical ways (e.g., being more uncomfortable with a transgender woman wearing makeup). Because these studies controlled for the actual stimuli being described as cisgender or transgender, the differences observed could only be explained by people's pre-existing ideas about transgender and cisgender people, rather than by any visual cues in the actual stimuli.

Despite this recent evidence that gender modality shapes certain aspects of gender-relevant categorization, we know of no work asking whether people perceive individuals as part of cohesive gender modality-based categories, nor work directly testing whether people automatically encode other people's gender modality. For example, do people mistake transgender people for one another more often than they mistake a transgender person for a cisgender person (or vice versa)? Despite the decades of use of the Who-Said-What task, we know of no uses of it to assess the presence (or absence) of automatic encoding along the dimension of gender modality. In addition, while some aspects of perceivers' identities have been linked to variation in the categorization and perception of transgender targets (e.g., political orientation, Stern & Rule, 2018; gender identity, Mao et al., 2019), we know of no work asking whether cisgender and transgender *perceivers* differ in their categorization of transgender and cisgender *targets*.

1.3. The role of familiarity/contact in categorization

Individuals' own experience with categories can sometimes impact their perception of categories and category members. A classic example is the cross-race face effect, in which people tend to better remember the faces of people who are in their own racial group than people in other racial groups (e.g., Malpass & Kravitz, 1969). Similar effects have been observed for other categories (e.g., age, Wright & Stroud, 2002; gender identity, Wright & Sladden, 2003). Suggesting that familiarity may play a role in this process, members of groups who have more experience

with outgroups than ingroups do not show this effect. For example, while East Asian adults who immigrated to Canada show the own-race face effect, East Asian adults born and raised in Canada do not (Zhou et al., 2019).

Experience plays a role in a variety of other facial perception tasks as well. For example, children who have experienced abuse—and therefore are presumably more familiar with displays of anger—have a lower threshold to identify a face as angry than children who have not been maltreated (Pollak & Kistler, 2002), and adults who are clinically depressed show better recall for previously seen sad faces, and worse recall at previously seen happy faces, compared to people who are not depressed (Ridout et al., 2003). These findings suggest that exposure to certain types of faces or categories of people may make one more or less attentive to, or have greater memory for, those faces.

However, some aspects of categorization appear less sensitive to varying experience. For example, even young children in a gender-neutral school automatically categorize others by gender identity, despite lower levels of gender stereotyping and more willingness to play with other-gender children (Shutts et al., 2017). In the present work, we explore the possible role of familiarity/contact on gender-relevant categorization and face perception by recruiting participants who vary on two dimensions expected to impact gender-relevant categorization: gender modality and age group. We also measure this variable to assess its relation to gender categorization within and across groups.

Differences in familiarity with transgender people might shape not only people's implicit gender categorization, but their explicit categorization of gender as well. These differences in experiences could lead people to hold different explicit views of gender identities, gender modalities, or both. For example, recent work with cisgender adults found that they tended to rely on identity among cisgender targets and modality among transgender targets (Gallagher & Bodenhausen, 2021), but how these judgments might vary across participant groups is unknown.

1.3.1. Possible variation by perceiver gender modality

There are several reasons why a perceiver's own gender modality (as transgender or cisgender) might impact their gender categorization. First, transgender people may be more familiar with (other) transgender people, given that people tend to show homophily—associating more often with people who share their identities (McPherson et al., 2001). For example, a recent study suggested that most trans people have at least one close friend who is transgender (Boyer & Galupo, 2018), while only about 40% of all Americans know at least one transgender person (Minkin & Brown, 2021). As a result of this (possibly) greater exposure to other transgender people, trans people may differ from cisgender people in their fluency or accuracy in categorizing transgender people in line with their gender identities. Said differently, cisgender people may be *impaired* at categorizing or recognizing individual transgender or other gender diverse people because of their relative lack of familiarity with the category. Greater familiarity and fluency in thinking about gender modality could also result in transgender people attending to this marker more than cisgender people, a possibility examined in the present work.

Another, not mutually exclusive, possibility is that transgender people may think about gender differently than cisgender people do. Some preliminary work suggests that transgender people may see gender as more fluid and less binary than cisgender people (e.g., Atwood et al., 2024; Fast & Olson, 2018), though not all work shows this pattern. For example, transgender and cisgender people may systematically vary in their explicit views of gender. Considering a social or medical transition might also lead some transgender people to reject more traditional notions of gender, to reflect more on who counts as a man or woman, or to think of gender as expanding beyond the categories of men and women (Serano, 2007).

In one recent study, cisgender, transgender, and nonbinary participants sorted morphed faces (i.e., faces that involved morphing a man's face and a woman's face to different degrees), one at a time, along a line from male to female (Atwood et al., 2024). Cisgender participants generally sorted the faces into two primary regions along the line, one for "men" and one for "women." Cisgender people were especially inaccurate in sorting the most androgynous faces (i.e., the faces that were an even mix of a man's and a woman's face). In contrast, the transgender and nonbinary participants were more accurate in sorting the androgynous faces, indicating that their representation of gender was less biased toward two discrete categories and likely more continuous. While this work explored a different aspect of gender categorization than the automatic encoding of gender identity and modality examined in the current work, it provides some evidence that cisgender and transgender/nonbinary people may differ on aspects of gender-relevant face categorization.

Much of the work comparing transgender and cisgender people's gender categorization has focused on young children. Some of this work supports the idea that group differences between transgender and cisgender people in gender categorization and thinking about gender more broadly may be present early in life. For example, 3- to 5-year-old transgender children (and to some extent their siblings) are more likely to believe that the gender of other people might change across development (e.g., a boy may later be a woman; Fast & Olson, 2018) compared to (unrelated) cisgender children. Six- to 8-year-old transgender children (and to some extent their siblings) endorse more acceptance of gender nonconformity in others than cisgender children (Olson & Enright, 2018). Finally, some work has suggested that elementary-aged transgender children are less likely to endorse gender stereotypes (Olson & Enright, 2018) or essentialize gender (Gülgöz et al., 2021) than cisgender children. These findings indicate that transgender people, beginning early in life, may have a more flexible and possibly more continuous representation of gender than cisgender people. However, whether these possible differences mean that they will also differ in their attention to gender categories is unclear.

Not all research leads to the conclusion that cisgender and transgender people will differ in terms of their gender categorization. There is also work suggesting a lack of difference between how cisgender and transgender children think about and use gender categories. For example, some other work on gender stereotyping (deMayo et al., 2021; Rubin et al., 2020) and gender essentialism (Gülgöz, DeMeules, et al., 2019) has observed no significant differences between these groups. A lack of significant difference between cisgender and transgender participants has also been observed in studies on children's gender-relevant preferences (e.g., preferences for gendered toys and clothing) and gender identity (e.g., Gülgöz, Glazier, et al., 2019; Olson et al., 2015).

Most relevant to the current work, we know of one study asking whether gender diverse (i.e., a mix of transgender and other gender nonconforming) 3- to 5-year-olds differed from cisgender 3- to 5-year-olds in their use of gender identity on a Who-Said-What task (Glazier et al., 2020). That study observed no significant difference between groups. However, even these results are complex to interpret upon closer look. Not only was the sample size small ($N = 71$ gender diverse children), but the authors reported (in the supplement) that when the comparison involved only the transgender ($N = 41$) and a matched set of cisgender ($N = 39$) children (i.e., excluded all other gender diverse children as in the original preregistration), the difference between groups was significant such that transgender children attended to gender more than the cisgender group. In sum, there are reasons to suggest that a perceiver's gender modality might shape their gender-relevant categorization, but the existing data are mixed and often involve small samples.

1.3.2. Possible age-related variation

There may also be a generational divide in how people are thinking about gender, especially gender modality. Evidence for this possible shift comes in a few forms. For example, age has been linked to views

Table 1
Study 1 Demographics ($N = 234$).

Age		M = 36.60 SD = 12.66 Range = 18–76
Political Ideology [*1/Very Liberal – 7/Very Conservative*]		M = 3.03 SD = 1.65
Bachelor's Degree or Higher		35%
Gender	Cisgender Men	43%
	Cisgender Women	57%
Race & Ethnicity	White or Caucasian, Non-Hispanic	65%
	White or Caucasian, Hispanic	8%
	Black or African, Non-Hispanic	8%
	Asian, Non-Hispanic	9%
	Multiple Options Chosen, Non-Hispanic	6%
	Multiple Options Chosen, Hispanic	1%
	Other Racial/Ethnic Group	2%

that someone's gender can differ from their sex (at least among Democrats) and support for transgender rights (Parker et al., 2022). Also, a higher percentage of current youth (i.e., adolescents and emerging adults) identify as transgender or nonbinary than mature adults. For example, 1.4% of 13–17 year olds and 1.3% of 18–24 year olds identify as transgender, but only 0.5% of 25–65 year olds do (Herman et al., 2022). This difference in identity has meant that *cisgender* people of different ages vary in their levels of exposure to transgender and nonbinary people, and age is now correlated with knowing someone who is transgender (Minkin & Brown, 2021). A Pew poll found that while 35% of Gen Zers know a person who "prefers to go by gender-neutral pronouns", only 16% of Gen Xers do (Parker & Igielnik, 2020). Insofar as familiarity with transgender people might play a role in the categorization and perception of trans people, we might observe differences by age. Further, age is related to holding different beliefs about gender, beliefs that might have implications for categorization. For example, while the majority of Gen Z'ers (59%) believe that a form or online profile should include more gender options than "man" and "woman", only a minority of Gen X'ers (40%) do (Parker & Igielnik, 2020).

2. The current studies

In the current studies, we examined implicit and explicit gender-relevant categorization in the domains of gender identity and gender modality. We first conducted a preliminary study (Study 1) of candidate measures to select those to include in our primary, registered study (Study 2). In Study 2, we aimed to compare gender-relevant categorization across four groups of participants who systematically vary in their experiences: transgender and cisgender youth (ages 12–22) and mature adults (ages 35–65).

We assessed implicit and explicit gender categorization. Implicit and explicit measures do not always show the same results (Kahneman, 2003; Sloman, 1996), including when people are reasoning about gender (Devine, 2001; Eidson & Coley, 2014; Greenwald & Banaji, 1995; Moss-Racusin et al., 2012). We therefore asked whether group differences appeared only on a more explicit measure of gender-relevant categorization, or on a more implicit measure as well. We also included two measures of possible differences between transgender and cisgender participants, and between youth and mature adults, that we thought might help explain any observed group differences: a measure of contact with transgender people, and a measure of beliefs about gender.

3. Study 1 (Non-registered Study)

Study 1 had three main goals: (1) to confirm that the tasks included in Study 2 were understandable to participants, (2) that the tasks produced the expected results in the sample that was easiest to recruit in Study 2, and (3) to provide an effect size estimate for Study 2, the registered study. All measures, manipulations, and data/participant exclusions are reported in the manuscript or its Supplementary Material.

3.1. Methods

De-identified data and all materials are available on OSF osf.io/pxfus/, including data from excluded participants.

3.1.1. Participants

We recruited 300 participants living in the United States on the Prolific online platform (Palan & Schitter, 2018). As this initial study was exploratory, we aimed for a relatively large sample to ensure adequate power. We excluded 12 participants for failing a Winograd schema question designed to detect bots and ensure English comprehension (Levesque et al., 2012) and 38 participants for failing a comprehension check (described below). We also excluded 16 participants who did not explicitly say they were not gender diverse, because these participants did not comprise a large enough sample to analyze by itself and we had theoretical reasons to suspect their responses may differ from the cisgender sample (as described above). We intentionally recruited a sample of transgender people for comparison in Study 2. Demographics of our final sample are included in Table 1.

3.1.2. Procedure

After completing an online consent process, participants were shown definitions of four gender categories: transgender women, transgender men, cisgender women, and cisgender men. For each category, they were told what the person's birth certificate said their sex was at birth, and how the person identifies now. We did this before participants began the task to avoid participant confusion about the terms used throughout the study (i.e., transgender and cisgender), and to ensure participants were participating in the task with at least the same minimum information. The labels for these categories were color-coded, with transgender labels framed in one color and cisgender labels framed in another color (purple and orange, randomized across participants). Participants then proceeded to complete five tasks, three of which are central to our analyses and also included in Study 2; these tasks are described here. The other two tasks were exploratory and not included

in Study 2; they are described in the Supplementary Material.

Who-Said-What Task. We drew on existing literature to design our Who-Said-What task (Castelli et al., 2022; Klauer & Wegener, 1998; S. E. Taylor et al., 1978). Participants saw three different "conversations." Each conversation included statements from eight different people: two transgender men, two transgender women, two cisgender men, and two cisgender women. Each face was presented with a statement (e.g., "I like long road trips"), and a gender group label (e.g., "Transgender Woman" presented above the face). The labels had the same-color frames as the definitions at the start of the study. This was done because we knew that gender identity was salient through other physical cues (e.g., physical appearance, hair styles), and so we wanted to ensure that the gender modality dimension was also highlighted in a salient way. Each conversation was about a single topic (vacation, animals, or food). As we did not find an existing Who-Said-What task that had three sets of eight statements, we generated them ourselves (they can be seen in the OSF materials).[2] After watching all three conversations, participants answered 12 test questions. In each test question, they were shown one phrase they had seen in the first part of the task and asked which person had said the phrase (one statement per gender category was shown from each conversation). Participants chose from among the 8 people who had been included in the conversation where the phrase appeared, and gender labels were again presented with the faces.

As well as differing in topic, each conversation differed by target race (Black, Asian, or White). While race was not a variable of interest in this work, we included people from three racial groups to improve stimulus sampling, to improve the generalizability of the findings, and to move away from field norms of using White faces as a default. By separating conversations by race, we minimized the possibility of interactive effects of target race with conversation. The orders of conversations, test trials, and response options, and the association between racial group and conversation topic, were all randomized. The within-conversation association between gender category and statement was quasi-randomized. We used 48 happy faces from the RADIATE database (8 men and 8 women from each racial group, Conley et al., 2018). For each racial group, each participant was randomly assigned to one of two mutually exclusive sets of 8 faces (4 men and 4 women), to again increase generalizability and improve stimulus sampling. Cisgender and transgender men were represented by men's faces, and cisgender and transgender women were represented by women's faces, with each face equally likely to represent a cisgender or transgender target. Participants saw two faces of each category (i.e., two transgender men, two cisgender women, and so on), with one of the of the two randomly selected to have their sentence presented in a test trial. This task took participants from 1.9 to 34.1 min (M = 5.6, SD = 3.8, Median = 4.6).

Continuous Similarity Judgments. Participants answered explicit questions about the similarity between the four gender categories by rating how similar two categories are to each other, for every possible category-pair (e.g., *How similar are Transgender Women and Cisgender Women?*) from 0/*Completely Different*-100/*Completely the Same* (based on Gallagher & Bodenhausen, 2021). Pairs were presented in a random order, and the order in which categories were mentioned in each pair was randomized.[3]

Comprehension Check. After completing the tasks, participants were asked two comprehension check questions. In each question, a celebrity who had changed their name was described. A brief life story was presented, which included their sex assigned at birth and their current gender identity. After reading the vignette, participants chose the gender category to which that person belonged. Participants read one vignette about a transgender celebrity (either a man or a woman) and one vignette about a cisgender celebrity (either a man or a woman[4]). 260 participants got both questions correct. Thirty-five participants failed only the transgender question, 3 participants failed only the cisgender question, and 2 participants failed both questions. Participants who got either question incorrect were excluded from our analyses (two participants had already been excluded for failing the Winograd questions).

Finally, participants answered a series of mostly-demographic questions. We also used this section to make sure that participants were living in the US (entering the two-letter code for their state of residence) and to exclude possible bots and ensure English comprehension (two Winograd Schema questions).

Table 2
Who-Said-What: Repeated-measures ANOVA Predicting Number of Errors (adjusted) by Error Type.

Effect	DF	F	p	partial η^2
Identity Error-Type	1, 229	273.74	< .001	0.54
Modality Error-Type	1, 229	58.24	< .001	0.20
Identity Error-Type x Modality Error-Type	1, 229	19.74	< .001	0.08

3.2. Results

All analyses were conducted in R, using the knitr, lme4, psych, rstatix, sjPlot, and tidyverse packages (Bates et al., 2015; Kassambara, 2022; Lüdecke et al., 2022; Revelle, 2019; Wickham et al., 2019; Xie, 2023). When possible, we report standardized effect sizes (e.g., Cohen's d). For multi-level regressions, we rely on point estimates instead – we consider them maximally informative because standardized effect size reporting in this context is still disputed (e.g., Muradoglu et al., 2023). Any instance where a participant did not answer a question is treated as a non-answer, not as an incorrect response. We present results for the Who-Said-What task and the Continuous Similarity Judgments in the main manuscript. We also tested for effects of participant gender identity; these analyses showed no significant ingroup effect (Wright & Sladden, 2003), only inconsistent and difficult-to-interpret interactive effects, so we report these analyses in the Supplementary Material. The Supplementary Material also include the results of two additional tasks that were completed by participants in Study 1 but not in Study 2.

3.2.1. Who-Said-What Task (Implicit Categorization)

On average, participants chose the correct answer 37% of the time (this differed significantly from the chance value of 12.50%; $t(233) = 19.11$, $p < .001$, $d = 1.25$). Our central interest was whether participant errors tended to be within-identity (i.e., confusing one woman for another), within-modality (i.e., confusing one transgender person for another), within-both (i.e., confusing one cisgender man for another), or between-both (i.e., confusing a cisgender woman with a transgender man). We tested this in a 2 × 2 repeated-measures ANOVA predicting the number of errors of each type made by each participant (this analysis omits data from all accurate trials, and from the four participants who

[2] In Study 1, one of the sentences used the word watermelon. In Study 2, to avoid racialized connotations of the fruit, we switched the word to cantaloupe.

[3] Due to a programming error in Study 1, participants were always asked about the similarity of Cisgender Women and Cisgender Men with the categories listed in that order (order of categories in pair was not randomized). This was corrected in Study 2.

[4] In Study 1, we made an error in having Miley Cyrus as our example of a cisgender woman who has changed her name – at the time, we were unaware that Miley Cyrus has publicly identified as genderfluid. In Study 2, we included Reese Witherspoon in our comprehension check instead. We retain the Study 1 comprehension check because the information we provided in the question matched our initial definitions of the gender groups, clearly defining Miley as assigned female at birth and currently using she/her pronouns (of all recruited participants who saw the Miley Cyrus comprehension check, 98.68% identified her as a cisgender woman). We regret the error.

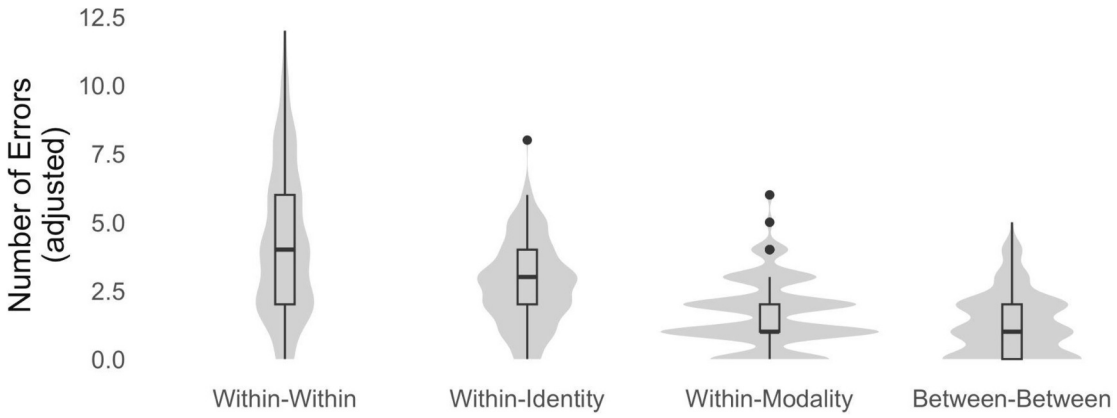

Fig. 1. Who-Said-What: Distribution of Number of Errors (adjusted).

were entirely accurate). On each trial, a participant could make only one wrong choice that was a within-both error (because the other within-both answer was correct), while they could make two wrong choices of each other kind. We account for this by multiplying the within-both error count by two before conducting the ANOVA, as is standard in analysis of the Who-Said-What paradigm (as discussed in Stangor et al., 1992; S. E. Taylor et al., 1978).

Within-identity errors were more frequent than cross-identity errors, within-modality errors were more frequent than cross-modality errors, and these two effects interacted such that within-both errors were most likely (see Table 2 and Fig. 1; $M_{\text{within-both}}$: 4.21 [SE = 0.18]; $M_{\text{within-identity}} = 2.71$ [SE = 0.10]; $M_{\text{within-modality}} = 1.62$ [SE = 0.08]; $M_{\text{between-both}} = 1.25$ [SE = 0.08]). This also means that errors that differed on gender modality and identity were particularly unlikely. To illustrate, for a transgender woman target, participants were most likely to mistake her statement as being stated by another transgender woman, then a cisgender woman, then a transgender man, and were least likely to mistake that statement as being said by a cisgender man.

As a follow-up test, we examined whether the target's gender group was related to the frequency of different kinds of errors. For instance, did participants make more within-modality errors when the target was transgender than when the target was cisgender? Did participants make more within-identity errors when the target was a man than when the target was a woman? We tested this in a 2 × 2 × 2×2 repeated-measures ANOVA (target identity, target modality, identity error-type, modality error-type). In this model, we saw the same pattern as in our 2×2 model. Target gender group had no significant effect on number of errors overall (*ps* > 0.23) and did not interact with error-type in predicting

Table 3
Who-Said-What: Number of Errors of Each Type (adjusted) Based on Target Gender Category.

Target Gender Category	Within-Both Errors	Within-Identity Errors	Within-Modality Errors	Between-Both Errors
Transgender Man	1.10 [0.09]	0.64 [0.05]	0.42 [0.04]	0.33 [0.03]
Transgender Woman	1.04 [0.09]	0.70 [0.05]	0.43 [0.04]	0.32 [0.04]
Cisgender Man	1.01 [0.09]	0.68 [0.05]	0.37 [0.04]	0.30 [0.03]
Cisgender Woman	1.05 [0.09]	0.68 [0.04]	0.39 [0.04]	0.30 [0.03]

Note: The mean number of errors of each type is displayed, with standard error displayed in brackets.

Table 4
Similarity Judgments: Multi-level Regression Predicting Ratings by Pair Type.

Predictors	Estimate	95% CI	T statistic	p
(Intercept)	−12.11	−14.44 - -9.78	−10.20	< .001
Between-Both vs Within-Either	5.53	2.96–8.10	4.23	< .001
Within-Identity vs Within-Modality	13.26	10.29–16.23	8.77	< .001
Between-Both: TM & CW vs TW & CM	−0.68	−4.88 - 3.51	−0.32	.749
Within-Identity: TM & CM vs TW & CW	−3.95	−8.15 - 0.24	−1.85	.065
Within-Modality: TM & TW vs. CM & CW	8.87	4.67–13.06	4.15	< .001
Random Effects				
σ^2	535.28			
$\tau_{subject}$	240.32			
ICC	0.31			
$N_{subject}$	234			
Observations	1404			
Marginal R^2 / Conditional R^2	0.054 / 0.347			

Note: This table shows the regression results, with one row showing the result of each orthogonal contrast code as described in the text. *TM = transgender men, CM = cisgender men, TW = transgender women, CW = cisgender women.* For each test, DF = 1396.

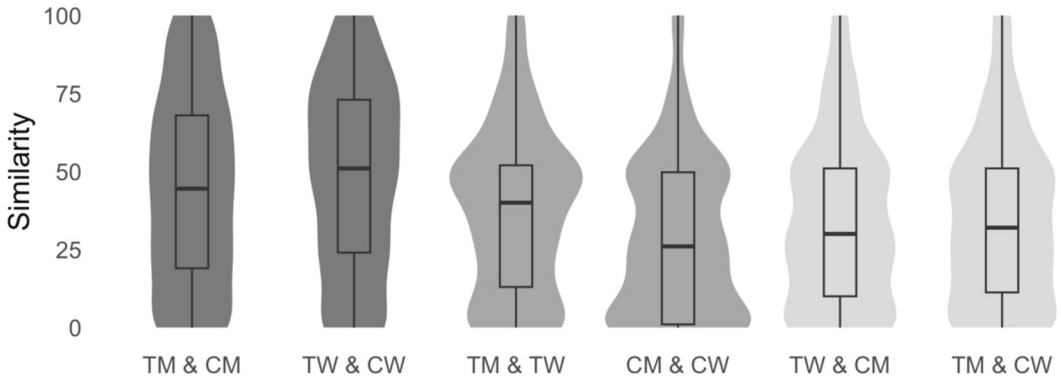

Fig. 2. Similarity Judgments: Distribution of Ratings.
Note: Within-identity pairs (dark gray) were judged as the most similar compared with within-modality (medium gray) and between-both (light gray) pairs. TM = transgender men, CM = cisgender men, TW = transgender women, CW = cisgender women.

number of errors ($ps > 0.47$; see Table 3 for mean numbers of errors; see Table S1 in Supplementary Material for full ANOVA results). That is, though participants *used* both target identity and target modality when attempting to recall Who-Said-What, they did this to the same extent regardless of the target's identity or modality.

3.2.2. Continuous Similarity Judgments (Explicit Categorization)

Our Similarity Judgment data is nested, with each participant (level-2) making six ratings (level-1) about different pairs of gender categories. To contrast the six ratings, we defined five orthogonal effect codes, making as many comparisons as possible without violating the required independence of each code (as described in Judd et al., 2017). Our first effect code compares ratings of pairs that share neither identity nor modality (between-both pairs: transgender women and cisgender men, transgender men and cisgender women) to ratings of pairs that share either identity or modality (within-modality pairs: cisgender men and cisgender women, transgender men and transgender women; within-identity pairs: cisgender women and transgender women, cisgender men and transgender men). Our second effect code compares ratings of within-identity pairs to ratings of within-modality pairs. Our third, fourth, and fifth effect codes contrast (respectively) the two ratings of between-both pairs, the two ratings of within-identity pairs, and the two ratings of within-modality pairs.

To account for the nested structure of the data, we could have used these effect codes either inside a multi-level regression or a 1 × 6 repeated-measures ANOVA. We opted for the multi-level regression because, along with the effects defined above, it provides an estimation of the model intercept, which we can interpret as whether participants were generally rating pairs of gender groups as similar (above 50) or dissimilar (below 50). We therefore conducted a multi-level regression with five fixed effects (the effect codes defined above), and a per-participant random intercept ($M_{TM\ \&\ CM} = 44.39$ [SE = 2.00], $M_{TW\ \&\ CW} = 48.34$ [SE = 1.98], $M_{TM\ \&\ TW} = 37.54$ [SE = 1.73], $M_{CM\ \&\ CW} = 28.67$ [SE = 1.67], $M_{TW\ \&\ CM} = 33.86$ [SE = 1.76], $M_{TM\ \&\ CW} = 34.54$ [SE = 1.75]; see Table 4 for regression results, and Fig. 2 for a visual representation of distributions).

We subtracted the scale midpoint (50) from each rating so that the intercept was calculated relative to the midpoint. Overall, participants rated all pairs as relatively dissimilar (i.e., below the scale midpoint, intercept $p < .001$). Similar to our Who-Said-What results, participants rated groups sharing either identity or modality as more similar than groups sharing neither, and they rated groups sharing identity as more similar than groups sharing modality ($ps < 0.001$). Similarity ratings for between-both pairs and within-identity pairs did not differ based on the specific pair ($ps > 0.06$). Similarity ratings for within-modality pairs, however, did differ by modality: Transgender categories were rated as more similar to each other than cisgender categories.

Given the significant difference between the both-transgender (TM & TW) pair and the both-cisgender (CM & CW) pair ratings, we further tested whether all non-shared-identity (the four pairs on the right of Fig. 2) pairs differed from each other using paired *t*-tests with Bonferroni corrections. The CM & CW pair was rated significantly less similar than the TM & TW pair (as also shown in the main model; $t(233) = -4.98$, $p < .001$, $d = -0.33$), significantly less similar than the TM & CW pair ($t(233) = -3.02$, $p = .017$, $d = -0.20$), and marginally less similar than the TW & CM pair ($t(233) = -2.58$, $p = .063$, $d = -0.17$). No other differences were significant (TM & TW vs TM & CW: $t(233) = 1.46$, $p = .879$, $d = 0.10$; TM & TW vs. TW & CM: $t(233) = 1.90$, $p = .349$, $d = 0.12$; TM & CW vs TW & CM: $t(233) = 0.47$, $p > .9$, $d = 0.03$).

3.2.3. Connecting implicit & explicit categorization

The results from the Who-Said-What task and the Similarity Judgments show somewhat similar patterns—in both tasks, participants recognized and used gender identity, gender modality, and the relationship between the two. However, we also wanted to test whether those overall patterns mapped onto individual differences in tendencies to rely on identity, modality, or their combination. To test this, we calculated the average similarity rating for each participant for each type of pair (within-identity, within-modality, between-both). From the Who-Said-What task, we calculated—again, for each participant—the number of errors they made that were of each type. In all three cases, the number of errors and explicit similarity ratings were not significantly correlated with one another (within-identity: $r(232) = -0.12$, $p = .068$; within-modality: $r(232) = 0.06$, $p = .338$; between-both: $r(232) = 0.06$, $p = .352$). Using a post-hoc power analysis, we note that we had 80% power to detect effects as small as $r = 0.18$.

3.3. Study 1 Discussion

Study 1 provides preliminary evidence that gender modality plays an important role in gender-relevant categorization. We replicated prior results from the Who-Said-What task showing that cisgender participants encoded target gender identity, and we expanded on this prior work to show that participants also encode target gender *modality*. To our knowledge, this is the first instance in which such an effect of modality has been demonstrated using this task, adding much-needed insight into how existing theories of gender categorization map onto to the increasingly varied landscape of gender identities. The results of Study 1 thus lay the groundwork for the planned systematic comparisons across age group and gender modality in Study 2.

On both the Who-Said-What task and Similarity Judgments, participants in Study 1 relied more heavily on gender identity than gender modality. However, the two tasks also showed important differences. On Similarity Judgments, participants rated transgender men and women as

more similar to each other than cisgender men and women. On the Who-Said-What task, participants' pattern of errors did not differ by target gender group – participants mis-remembered cisgender women as cisgender men (and vice versa) just as often as they mis-remembered transgender women as transgender men (and vice versa). Moreover, despite similar main effects in the two tasks, we did not find between-task correlation—the group-level findings did not reflect patterns of individual differences. Including both our proposed implicit and explicit measures of gender-relevant categorization in Study 2 can therefore offer meaningful information about how gender identity and gender modality shape gender-relevant categorization processes. In Study 2, we tested whether these implicit and explicit effects differ by the experiences and identities of transgender and cisgender youth and mature adults.

4. Study 2 (Registered Study)

In Study 2, we asked whether groups that systematically differ in their experiences with gender modality differ on implicit and explicit measures of gender-relevant categorization. Specifically, we asked whether cisgender and transgender youth and mature adults differ in the degree to which they automatically encode gender identity and gender modality, and the degree to which they explicitly judge similarity between gender groups. Further, we replicated the finding from Study 1 that people automatically encode gender modality, at least when it is made salient. Finally, we tested whether participant group differences in explicit or implicit categorization could be explained by familiarity/contact with transgender people and/or beliefs that gender is fluid.

Study 2 is the registered study in this manuscript. All materials, including power analyses, data, and analysis scripts, can be seen here: osf.io/qb6mj. The in-principle-accepted registered report can be seen here: osf.io/yv9gk. In any instances where we have deviated from the Stage 1 Registered Report, or encountered decisions not addressed in the preregistration, we have noted this in the text.

4.1. Method

4.1.1. Participants

For this study, we recruited participants from four different groups: transgender youth, cisgender youth, transgender adults, and cisgender adults,[5] requiring at least 100 participants per group to proceed to analysis. We recruited the youth sample (age 12–22) between September and November 2023 via the annual survey of a longitudinal project about gender development in gender diverse youth and their cisgender peers (see Olson & Gülgöz, 2018, Gülgöz et al., 2019, and Durwood et al., 2024 for details). At the end of each month, we determined how many youth were eligible for inclusion in this manuscript (see Supplementary Material), and recruited that many participants via Prolific for our corresponding mature adult samples (age 35–65; recruited between October and December 2023; participants from Study 1 were not able to participate). By doing this monthly, we ensured that the samples participated at similar times. Doing so was crucial, given how political transgender issues have become in the U.S. context and therefore how participant attitudes about trans people could potentially shift rapidly with changing national events.

Youth were considered transgender if they said "yes" to the survey question *Are you currently transgender, nonbinary, or gender diverse?*, and cisgender if they said "no" to that question. Adults were considered transgender if they said "yes" to the Prolific demographics question *Does your current gender differ from the one you were assigned at birth?*, and cisgender if they said "no" to that question. Although we acknowledge that not all participants in these transgender samples necessarily use the label "transgender" to describe themselves, we describe the samples in this way to capture the experience they share (and do not share with the cisgender sample) of identifying with and/or enacting a gender category that does not align with the one they were assigned at birth.[6] After exclusions (see Supplementary Material for details), our main sample included 264 transgender youth, 249 cisgender youth, 148 transgender adults, and 249 cisgender adults (see Table 5 for demographic information). We additionally have data from a sample of 75 cisgender siblings of gender diverse youth, analyzed separately in the Supplementary Material.

4.1.2. Procedure

The exact procedure differed slightly between our longitudinal youth sample and our adult online sample. Adult participants only completed measures for this study, and all measures, manipulations, and data/participant exclusions are reported. For the youth sample, these tasks were embedded in a battery of measures that are also part of a larger longitudinal research project. A small number of demographic questions preceded the measures for the current study (see Supplementary Material for question text); measures presented after those for this study are not reported in this manuscript. These include measures assessing topics like mental health, medical care, and sexuality.

All participants completed the tasks for this study in the following order: the Who-Said-What task and Continuous Similarity Judgments as described in Study 1, a series of questions about different identity domains (e.g., *Are you Asian? Are you transgender?*; this latter measure was initially intended to classify participants as transgender or cisgender, but is not used in the final manuscript; see Supplementary Material for details), and two potential mediators (randomized order): Familiarity/Contact with Transgender People and Openness Towards Non-Binary Gender (see below). For the Prolific samples (as in Study 1, but not with the longitudinal sample), gender categories were defined at the beginning of the study, the comprehension check was included between the main tasks and the demographics section, and an English fluency check was embedded in demographics at the end of the study.

Potential Mediator 1: Familiarity/Contact with Transgender People. We assessed quality and quantity of contact with transgender people via two measures used in a recent paper by Fine et al. (2023), modifying items to refer to "transgender people" instead of "GNC [gender nonconforming] individuals." To assess quality of contact, we used three items (modified from Kteily et al., 2019; Voci & Hewstone, 2003) that assess how *pleasant*, *cooperative*, and *superficial or insincere* [reverse coded] participants' interactions with transgender people are (each rated on a 7-point scale from *1/Not at all – 7/Very much so*; $\alpha = 0.77$). Quantity of contact was assessed with four questions (again, modified from Voci & Hewstone, 2003): quantity of contact (*1/No*

[5] In this design, age group is aligned with recruitment venue – youth were part of the longitudinal cohort, while mature adults were recruited via Prolific. To see whether recruitment venue alone was related to task performance, we recruited two additional samples via Prolific – 100 cisgender young adults and 100 transgender young adults. In short, the findings of these young adults generally mirror the longitudinal youth sample, though the cisgender young adults from Prolific endorsed similarity of within-identity pairs (e.g., cisgender and transgender men) less than cisgender youth from the longitudinal sample. For details, see the Supplementary Material.

[6] Due to recruitment challenges, we updated our criteria to classify participant modality and age group during recruitment (before beginning any analysis). This was approved by the editor and is described in full in the Supplementary Material. Replicating the analysis of the Who-Said-What task only with participants who would have qualified under our original criteria shows the same general pattern of means, though the group differences are not significant (potentially owing to the very small sample size for transgender adults, $N = 37$; see results in Supplementary Material). Replicating the analysis of the Similarity Judgments task only with participants who would have qualified under our original criteria shows the same pattern of results as that presented in the main manuscript (see results in Supplementary Material).

Table 5
Study 2 Demographics.

		Transgender Youth	Cisgender Youth	Transgender Adult	Cisgender Adult
Total N		264	249	148	249
Age		M = 15.06 SD = 2.25 Range = 12–22	M = 14.90 SD = 2.36 Range = 12–21	M = 42.81 SD = 7.83 Range = 35–65	M = 47.61 SD = 8.38 Range = 35–65
Political Ideology [*1/Very Liberal – 7/Very Conservative*]		–	–	M = 1.93 SD = 1.34	M = 3.20 SD = 1.74
Bachelor's Degree or Higher		–	–	39%	37%
Gender	Boy/Man	24%	37%	23%	38%
	Girl/Woman	49%	63%	25%	62%
	Genderqueer	–	–	9%	0%
	Non-Binary	11%	0%	31%	0%
	Not Listed	12%	0%	11%	0%
	Prefer Not to Answer	0%	1%	–	–
	I don't know	4%	0%	–	–
Race & Ethnicity	White or Caucasian, non-Hispanic	67%	67%	71%	82%
	Multiracial, non-Hispanic	14%	18%	7%	4%
	Other Racial/Ethnic Group	9%	9%	17%	12%
	Unknown	2%	2%	0%	0%

Note: In some cases demographics are defined differently by sample – either because slightly different questions were asked of the longitudinal and Prolific samples (e.g., specific gender options), or because some demographic questions make more sense for adults than youth (e.g., education level). A – in the table represents that the option or question was not presented to all participants in that subset. For longitudinal participants, race/ethnicity is defined from the first time the participant ever reported their own race/ethnicity, and age is reported based on comparing the date of participation with the youth's birthdate. For participants recruited via Prolific, race/ethnicity and age are defined from the participant's response during this particular survey. Race/ethnicity groups appear in the table if at least 10% of one sample reported that race/ethnicity.

contact at all – 7/A great deal of contact), frequency of contact (*1/Never – 7/Very frequently*), number of transgender people known (*1/none – 11/ten or more*) and number of hours a week they spend with transgender people (*1/0 hours – 11/ten or more hours*). Responses across the four quantity items were z-scored and then combined ($\alpha = 0.92$). Further, we z-scored the average response to the quality items and then averaged the quality and quantity scores to produce a continuous measure of quality and quantity of contact with transgender people ($r(908) = 0.35$, $p < .001$). *Z*-scoring was done with respect to Study 2's main sample.

Potential Mediator 2: Openness to Non-Binary Gender. To assess beliefs about gender, we used two subscales from Molin et al.'s (2021) Openness Towards Non-Binary Gender scale: the 6-item Gender Fluidity subscale (e.g., *A person's gender can change over the course of their life*) and the 6-item Gender Categories subscale (e.g., *There are more than two gender categories*). Participants rated each item on a scale from *1/Completely Disagree – 6/Completely Agree* ($\alpha = 0.95$).[7] In the Stage 1 Registered Report, we did not explicitly state how we would combine items. We used mean response across answered items, and (to be consistent with the first potential mediator), we Z-scored this with respect to Study 2's main sample.

4.2. Results

Overview of Results. As expected, we fully replicated our Study 1 findings among the new sample of 249 cisgender adults (see Supplementary Material for details). Below, we report planned analyses of group comparisons for the Who-Said-What task and Similarity Judgments. In short, we found that Who-Said-What task performance among cisgender adults generalized to our other three participant groups (with a few small differences). Similarity Judgments showed that cisgender participants – and especially cisgender adults – saw within-identity pairs (e.g., transgender and cisgender women) as less similar than did transgender participants. Cisgender adults also saw within-modality pairs (e.g., transgender women and men) as less similar than did other participants, while transgender youth saw between-both pairs (e.g., transgender women and cisgender men) as less similar than did cisgender participants. Implicit and explicit task performance were largely uncorrelated. Participant groups differed on both potential mediators – Familiarity/Contact with Transgender People and Openness to Non-Binary Gender – and both of these mediated some but not all participant group differences in our two central tasks.

Analytic Approach. All Study 2 analyses were planned and preregistered as part of the Stage 1 Registered Report, unless noted otherwise. During Stage 1, we also conducted sensitivity analyses with our conservative minimum *N* of 100 per group. Since our final sample was more than twice that size, substantively increasing our statistical power, we have moved those analyses to the Supplementary Material. We do note that we observed statistically significant small-to-medium effects (e.g., *partial* $\eta^2 = 0.01$, $d = 0.30$, $r = 0.11$) across different analyses. We include all tests described in the Stage 1 Registered Report in the manuscript or Supplementary Material. As we did not register follow-up tests for significant participant group interactions in our main models, when interactions appeared, we conducted all six between-group comparisons using *t*-tests with a Bonferroni correction.

All analyses were conducted in R, using the emmeans, kableExtra, knitr, lme4, mediation, psych, rstatix, sjPlot, and tidyverse packages (Bates et al., 2015; Kassambara, 2022; Lenth, 2021; Lüdecke et al., 2022; Revelle, 2019; Tingley et al., 2014; Wickham et al., 2019; Xie, 2023; Zhu, 2024). We used the same analytic practices as described in Study 1. We used contrast codes to define participant gender modality (transgender = −0.5, cisgender = 0.5) and participant age group (youth = −0.5, adult = 0.5).

4.2.1. Who-Said-What Task (Implicit Categorization)

On average, participants chose the correct answer 36% of the time (SE = 1%; as predicted, this was higher than the chance value of 12.50%; $t(909) = 36.45$, $p < .001$, Cohen's $d = 1.21$). We used a 2 × 2 × 2 × 2 mixed ANOVA to predict the number of errors of each type made by each participant (using the same adjustment as in Study 1; this excludes 6

[7] In the Stage 1 Registered Report, we said that we would use a 1–7 response scale for this measure. Due to a programming error, we used to use a 1–6 scale instead. Nothing else about the measure changed.

Table 6
Who-Said-What: Mixed ANOVA Predicting Number of Errors (adjusted) by Error Type, Participant Group, and Interactions.

	Effect	DF	F	p	partial η^2
Within-Participant Effects of Error Type	Identity Error-Type	1, 900	847.65	< .001	0.48
	Modality Error-Type	1, 900	93.25	< .001	0.09
	Identity Error-Type x Modality Error-Type	1, 900	54.92	< .001	0.06
Between-Participant Effects of Participant Group	Participant Modality	1, 900	5.51	.019	0.01
	Participant Age	1, 900	0.45	.502	0.00
	Participant Modality x Participant Age	1, 900	4.28	.039	0.01
Within-Between Interactions	Participant Modality x Identity Error-Type	1, 900	0.09	.770	0.00
	Participant Modality x Modality Error-Type	1, 900	0.06	.803	0.00
	Participant Modality x Identity Error-Type x Modality Error-Type	1, 900	1.09	.296	0.00
	Participant Age x Identity Error-Type	1, 900	4.54	.033	0.00
	Participant Age x Modality Error-Type	1, 900	0.32	.572	0.00
	Participant Age x Identity Error-Type x Modality Error-Type	1, 900	1.10	.294	0.00
	Participant Age x Participant Modality x Identity Error-Type	1, 900	0.05	.827	0.00
	Participant Age x Participant Modality x Modality Error-Type	1, 900	0.93	.335	0.00
	Participant Age x Participant Modality x Identity Error-Type x Modality Error-Type	1, 900	1.27	.260	0.00

participants who answered every question correctly). We included participant group as between-participant factors (age group, gender modality, and their interaction), error type as within-participant factors (identity, gender modality, and their interaction), and the interactions of these as mixed factors (see Table 6 for ANOVA results, Table 7 for by-group means).

We observed the same within-participant effects as in Study 1, as well as two between-participant effects of group on number of errors: a significant effect of modality and a modality by age group interaction (there was no main effect of age group). Transgender adults were more accurate overall than cisgender adults ($t(284.99) = 2.80$, $p = .032$, $d = 0.30$), while both youth groups did not differ significantly from each other or either adult group (CY vs. TY: $t(508.14) = 0.22$, $p > .9$, $d = 0.02$; TA vs. TY: $t(283.97) = -1.75$, $p = .489$, $d = -0.18$; CA vs TY: $t(504.72) = 1.32$, $p > .9$, $d = 0.12$; CY vs. TA: $t(289.91) = 1.91$, $p = .346$, $d = 0.20$; CA vs. CY: $t(491.95) = 1.08$, $p > .9$, $d = 0.10$).

Finally, we observed one instance when a within-participant effect differed by participant group: the Who-Said-What identity effect (the tendency to make more within-identity than between-identity errors) was stronger among adults than youth. We followed up with a non-registered 2 × 2 × 2 mixed ANOVA only among youth (identity error-type x modality error-type x participant modality) to confirm that youth still showed the effect ($p < .001$; see Supplementary Material Table S11 for full ANOVA table), just less than the adults.

In sum, all groups most often confused targets who shared both gender identity and modality (e.g., confusing two transgender women), suggesting they processed both dimensions. The second most common confusion in all groups was confusing individuals who shared gender identity only (e.g. confusing a transgender woman for a cisgender woman). The participant group differences observed were that transgender adults made fewer overall errors than cisgender adults and that youth automatically encoded gender identity less than adults.

Table 7
Who-Said-What: Per-Participant-Group Number of Errors of Each Type (adjusted).

Participant Group	Within-Both Errors	Within-Identity Errors	Within-Modality Errors	Between-Both Errors
Transgender Youth	3.88 [0.17]	2.65 [0.09]	1.60 [0.08]	1.50 [0.08]
Cisgender Youth	3.70 [0.16]	2.92 [0.10]	1.65 [0.08]	1.41 [0.08]
Transgender Adults	3.86 [0.21]	2.69 [0.14]	1.29 [0.10]	1.23 [0.10]
Cisgender Adults	4.14 [0.17]	2.87 [0.10]	1.55 [0.09]	1.40 [0.08]

Note: The mean number of errors of each type is displayed, with standard error displayed in brackets.

Table 8
Similarity Judgments: Participant Group Differences in Within-Gender-Identity Ratings.

Predictor	Estimate	p
Intercept	61.48 [59.77–63.18]	< .001
Modality Group	−8.96 [−12.37 - -5.54]	< .001
Age Group	−18.40 [−21.81 - -14.99]	< .001
Modality Group x Age Group	−16.13 [−22.95 - -9.31]	< .001
Model Fit Statistics		
R^2 / R^2 adjusted	0.161 / 0.159	

Table 9
Similarity Judgments: Multi-level Regression Predicting Ratings by Participant Group, Rating Type, and Interactions.

	Predictors	Estimate	95% CI	T statistic	p
Within-Participant Effects	(Intercept)	−5.31	−6.39–4.23	−9.62	< .001
	Between-Both vs Within-Either	17.83	16.48–19.17	25.92	< .001
	Within-Identity vs Within-Modality	21.70	20.14–23.25	27.32	< .001
	Between-Both: TM & CW vs TW & CM	−1.90	−4.10–0.30	−1.69	.091
	Within-Identity: TM & CM vs TW & CW	−2.30	−4.51 - -0.10	−2.05	.040
	Within-Modality: TM & TW vs. CM & CW	6.68	4.47–8.88	5.94	< .001
Between-Participant Effects	Participant Modality	−5.22	−7.39 - -3.06	−4.73	< .001
	Participant Age	−3.45	−5.62 - -1.29	−3.12	.002
	Participant Modality x Participant Age	−11.24	−15.56 - -6.91	−5.09	< .001
Within-Between Interactions	Participant Modality x Between-Both vs Within-Either	−16.06	−18.75 - -13.36	−11.67	< .001
	Participant Modality x Within-Identity vs Within-Modality	−15.65	−18.76 - -12.54	−9.85	< .001
	Participant Modality x Between-Both: TM & CW vs TW & CM	−0.52	−4.92–3.89	−0.23	.818
	Participant Modality x Within-Identity: TM & CM vs TW & CW	1.20	−3.21–5.60	0.53	.595
	Participant Modality x Within-Modality: TM & TW vs. CM & CW	−2.39	−6.79–2.01	−1.06	.287
	Participant Age x Between-Both vs Within-Either	−9.02	−11.71 - -6.32	−6.56	< .001
	Participant Age x Within-Identity vs Within-Modality	−5.00	−8.11 - -1.88	−3.15	< .001
	Participant Age x Between-Both: TM & CW vs TW & CM	1.14	−3.26–5.54	0.51	.612
	Participant Age x Within-Identity: TM & CM vs TW & CW	−1.36	−5.76–3.04	−0.61	.545
	Participant Age x Within-Modality: TM & TW vs. CM & CW	2.97	−1.44–7.37	1.32	.187
	Participant Modality x Participant Age x Between-Both vs Within-Either	−7.16	−12.55 - -1.76	−2.60	.009
	Participant Modality x Participant Age x Within-Identity vs Within-Modality	−5.02	−11.25–1.21	−1.58	.114
	Participant Modality x Participant Age x Between-Both: TM & CW vs TW & CM	1.99	−6.82–10.80	0.44	.658
	Participant Modality x Participant Age x Within-Identity: TM & CM vs TW & CW	3.27	−5.54–12.07	0.73	.467
	Participant Modality x Participant Age x Within-Modality: TM & TW vs. CM & CW	−1.42	−10.23–7.39	−0.32	.752
	Random Effects				
	σ^2	543.31			
	$\tau_{subject}$	171.99			
	ICC	0.24			
	$N_{subject}$	910			
	Observations	5460			
	Marginal R^2 / Conditional R^2	0.216 / 0.404			

Note: This table shows the regression results. *TM = transgender men, CM = cisgender men, TW = transgender women, CW = cisgender women.* For each test, DF = 5434.

4.2.2. Similarity Judgments (Explicit Categorization)

We had a priori predictions about group differences in similarity ratings of within-identity pairs (e.g., cisgender and transgender men); we describe this analysis first, before proceeding to analysis of all six ratings. As predicted in our preregistration, youth rated within-identity pairs (the mean of the rating of transgender and cisgender men and the rating of transgender and cisgender women) as more similar than adults, and transgender participants rated these pairs as more similar than cisgender participants (see Table 8 for regression results). We further observed a significant interaction between age group and modality (we had no a priori prediction about the presence or absence of this interaction). Follow-up tests showed that cisgender adults rated these pairs less similar than all other participant groups (CA vs. TA: $t(372.26) = -10.08, p < .001, d = -1.01$; CA vs. CY: $t(487.05) = -6.85, p < .001, d$

Table 10
Similarity Judgments: Per-Participant-Group Ratings of Each Gender Category Pair.

Participant Group	TM & CM	TW & CW	TM & TW	CM & CW	TW & CM	TM & CW
Transgender	70.42	71.83	43.48	37.34	26.31	28.02
Youth	[1.53]	[1.50]	[1.64]	[1.72]	[1.61]	[1.60]
Cisgender	59.87	61.71	45.34	40.99	34.27	37.49
Youth	[1.73]	[1.75]	[1.76]	[1.68]	[1.58]	[1.67]
Transgender	68.03	72.43	46.81	37.10	32.18	33.74
Adults	[2.11]	[1.84]	[2.20]	[2.13]	[2.23]	[2.16]
Cisgender	42.98	44.55	36.96	30.35	34.67	35.76
Adults	[1.93]	[1.96]	[1.72]	[1.67]	[1.69]	[1.67]

Note: The mean similarity rating of each pair is displayed, with standard error displayed in brackets.

$= -0.61$; CA vs. TY: $t(464.40) = -11.72$, $p < .001$, $d = -1.04$), followed by cisgender youth (CY vs. TA: $t(343.40) = -3.84$, $p = .001$, $d = -0.39$; CY vs. TY: $t(493.68) = -4.81$, $p < .001$, $d = -0.43$), while transgender youth and adults rated these pairs as most similar (transgender youth and adults did not differ; $t(307.22) = -0.39$, $p > .9$, $d = -0.04$).

To analyze the full set of six ratings (not just those about which we had specific hypotheses), we used the same model structure as in Study 1, adding fixed effects for participant group (age group, modality group, and their interaction) and the cross-level interactions of participant group with our contrast codes. We again included a random intercept for participants to account for individual-level tendencies to rate pairs higher or lower on the scale (see Table 9 for regression results, Table 10 for means).

Across groups, the within-participant effects were the same as in Study 1, with one exception: transgender and cisgender women were rated more similar to each other than transgender and cisgender men (in other analyses, these ratings did not differ significantly). We also observed between-participant effects of participant group on overall similarity ratings. We followed up on this by comparing mean similarity ratings (across all six ratings) between the four groups. Transgender adults, cisgender youth, and transgender youth did not differ (CY vs. TA:

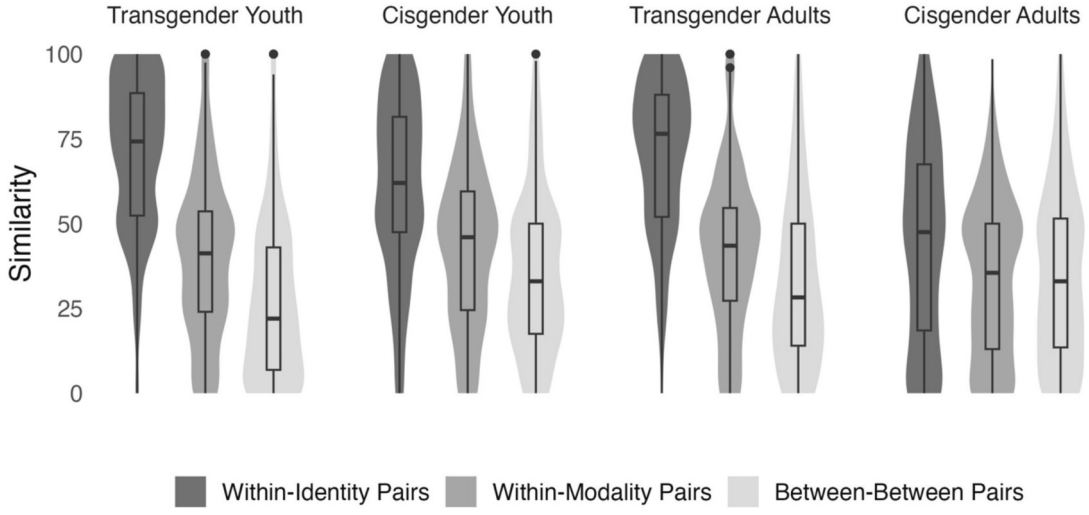

Fig. 3. Similarity Judgments: Per-Participant-Group Ratings of Within-Identity, Within-Modality, and Between-Between Pairs.
Note: This figure shows each participant's mean rating of the within-identity pairs, the within-modality pairs, and the between-between pairs.

Table 11
Who-Said-What & Similarity Judgments: Memory Errors (adjusted) and Participant Group Predicting Similarity Ratings.

	Mean Similarity Ratings					
	Within-Identity (Between-Modality)		Within-Modality (Between-Identity)		Between-Both	
Predictor	Estimate	p	Estimate	p	Estimate	p
Intercept	61.15 [57.78–64.53]	< .001	39.51 [37.11–41.92]	< .001	32.73 [30.32–35.15]	< .001
WSW Errors	0.12 [−0.94–1.18]	.826	0.07 [−1.16–1.31]	.909	0.03 [−1.28–1.33]	.965
Modality Group	−18.60 [−25.35 - -11.86]	< .001	−8.07 [−12.89 - -3.26]	.001	2.50 [−2.33–7.32]	.310
Age Group	−5.29 [−12.04–1.46]	.124	−3.08 [−7.90–1.73]	.209	2.61 [−2.22–7.43]	.289
Modality Group x Age Group	−19.56 [−33.05 - -6.06]	.005	−9.87 [−19.49 - -0.24]	.045	−6.57 [−16.22–3.08]	.182
WSW Errors x Modality Group	0.05 [−2.06–2.17]	.962	3.53 [1.06–6.00]	.005	2.18 [−0.43–4.79]	.101
WSW Errors x Age Group	−1.35 [−3.47–0.76]	.210	−0.67 [−3.14–1.80]	.594	−0.13 [−2.74–2.48]	.924
WSW Errors x Modality Group x Age Group	1.35 [−2.88–5.58]	.531	−0.26 [−5.20–4.68]	.918	0.32 [−4.90–5.54]	.905
Model Fit Statistics						
R^2 / R^2 adjusted	0.163 / 0.157		0.034 / 0.027		0.025 / 0.017	

Note: We display the overall effects of group here to display the full regression model; however, these effects are consistent with the more thorough description of group differences in Similarity Judgments described in the prior section, and we therefore do not describe them in detail here.

Table 12
Potential Mediators: Participant Group Differences.

	Familiarity with Transgender People		Openness to Non-Binary Gender	
Predictor	Estimate	p	Estimate	p
Intercept	0.03 [−0.02–0.08]	.230	0.04 [−0.02–0.10]	.222
Age Group	−0.11 [−0.21 - -0.01]	.036	−0.39 [−0.51 - -0.27]	< .001
Modality Group	−0.43 [−0.53 - -0.33]	< .001	−0.71 [−0.83 - -0.59]	< .001
Modality Group x Age Group	−0.54 [−0.74 - -0.34]	< .001	−0.88 [−1.12 - -0.65]	< .001
Model Fit Statistics				
R^2 / R^2 adjusted	0.102 / 0.099		0.216 / 0.213	

Table 13
Potential Mediators: Per-Participant-Group Means.

	Familiarity with Transgender People	Openness to Non-Binary Gender
Transgender Youth	0.16 [0.05]	0.36 [0.04]
Cisgender Youth	0.00 [0.05]	0.10 [0.06]
Transgender Adults	0.32 [0.06]	0.42 [0.06]
Cisgender Adults	−0.37 [0.05]	−0.73 [0.07]

Note: Brackets display standard errors. These scales were z-score with respect to the main sample.

$t(298.64) = -1.06$, $p > .9$, $d = -0.11$; CY vs. TY: $t(509.68) = 0.29$, $p > .9$, $d = 0.03$; TA vs. TY: $t(295.80) = 1.31$, $p > .9$, $d = 0.14$); while cisgender adults rated gender groups as overall less similar than the three other participant groups (CA vs. TA: $t(322.32) = -6.28$, $p < .001$, $d = -0.65$; CA vs. CY: $t(491.49) = -6.16$, $p < .001$, $d = -0.55$; CA vs. TY: $t(500.26) = -5.95$, $p < .001$, $d = -0.53$).

Finally, the relative ratings of within-identity, within-modality, and between-both pairs differed by participant group (visible in interactions between participant group and our first two contrasts; see Table 9 and Fig. 3). To follow up on this, we calculated per-participant mean ratings of within-modality and between-both pairs (the same way we handled within-identity pairs, above).[8] For within-modality pairs, cisgender adults rated these pairs as less similar than the other three participant groups (CA vs. TA: $t(304.56) = -3.41$, $p = .004$, $d = -0.36$; CA vs. CY: $t(495.62) = -4.52$, $p < .001$, $d = -0.40$; CA vs. TY: $t(510.88) = -3.21$, $p = .009$, $d = -0.28$), which did not differ from each other (CY vs. TA: $t(311.35) = 0.49$, $p > .9$, $d = 0.05$; CY vs. TY: $t(510.06) = 1.32$, $p > .9$, $d = 0.12$; TA vs. TY: $t(311.05) = 0.65$, $p > .9$, $d = 0.07$). For between-both pairs, transgender youth rated these pairs significantly less similar than cisgender adults ($t(506.26) = 3.67$, $p = .002$, $d = 0.32$) and cisgender youth ($t(510.44) = 4.11$, $p < .001$, $d = 0.36$), but did not differ from transgender adults ($t(293.77) = 2.26$, $p = .149$, $d = 0.23$). Cisgender adults, cisgender youth, and transgender adults did not differ from each other (CA vs. TA: $t(307.88) = 0.86$, $p > .9$, $d = 0.09$; CA vs. CY: $t(494.00) = -0.31$, $p > .9$, $d = -0.03$; CY vs. TA: $t(292.41) = 1.14$, $p > .9$, $d = 0.12$).

4.2.3. Connecting implicit & explicit categorization

In Study 1, we did not see evidence that Who-Said-What task performance and Similarity Judgments were significantly correlated. With Study 2's full data set, we tested whether this implicit-explicit relationship differed by age group or gender modality, using simultaneous regressions (see Table 11).[9] Across participant groups, we again observed no relationship between Who-Said-What errors and similarity ratings. In just one case – within-modality ratings and within-modality errors – this depended on participant modality. We calculated this correlation separately among cisgender and transgender participants (using a Bonferroni correction on the *p*-values), and saw that the number of within-modality Who-Said-What errors was positively associated with ratings of within-modality gender pairs among cisgender participants ($r(496) = 0.11$, $p = .024$), but not for transgender participants ($r(410) = -0.09$, $p = .152$). That is, cisgender participants who misattributed statements according to modality (e.g., attributing a statement by a transgender man to a transgender woman; attributing a statement by a cisgender woman to a cisgender man) also judged within-modality pairs (i.e., transgender men and women, cisgender men and women) as more similar.

4.2.4. Potential mediators: group-based differences in experiences and attitudes

As expected, youth participants (compared to adult participants) and transgender participants (compared to cisgender participants) reported greater familiarity with transgender people and more openness to non-binary gender (see Table 12 for regression results, Table 13 for means). Though we did not have a priori hypotheses about interactive age group by participant modality effects, we tested for and observed them for both mediators. Follow-up tests revealed that cisgender adults were less familiar with transgender people and had lower openness to non-binary identity than all other groups (Familiarity – CA vs. TA: $t(311.44) = -9.35$, $p < .001$, $d = -0.97$; CA vs. CY: $t(495.41) = -5.71$, $p < .001$, $d = -0.51$; CA vs. TY: $t(510.75) = -8.04$, $p < .001$, $d = -0.71$; Openness – CA vs. TA: $t(393.10) = -12.17$, $p < .001$, $d = -1.19$; CA vs. CY: $t(464.55) = -9.03$, $p < .001$, $d = -0.81$; CA vs. TY: $t(389.58) = -13.14$, $p < .001$, $d = -1.17$), while transgender adults and youth did not differ from each other (Familiarity: $t(328.16) = 2.14$, $p = .200$, $d = 0.22$; Openness: $t(278.03) = 0.75$, $p > .9$, $d = 0.08$). Cisgender youth reported significantly lower familiarity with transgender people than transgender adults ($t(320.02) = -4.25$, $p < .001$, d = −0.44), and did not differ from transgender youth ($t(510.92) = -2.34$, $p = .118$, $d = -0.21$). Cisgender youth reported significantly lower openness to non-binary gender than either transgender group (CY vs. TA: $t(353.53) = -3.91$, $p = .001$, $d = -0.40$; CY vs. TY: $t(457.82) = -3.86$, $p = .001$, $d = -0.34$).

4.2.5. Mediation analyses

We next conducted a secondary analysis, testing whether the group differences in our potential mediators could explain observed group differences in the Who-Said-What and Similarity Judgment tasks (please note, this is statistical, not causal mediation). For each participant group effect in our main tasks, we conducted a mediation analysis in three steps (see Table 14; this approach is simplified from our originally planned multi-level mediations, see Supplementary Material for more information). First, we arithmetically calculated a participant-level dependent variable to represent the metric on which the groups differed. Second, we defined our relevant participant groups, including which one was the reference group in the mediation. Finally, we conducted mediation analyses, using quasi-Bayesian confidence intervals as implemented in the *mediation* package (Tingley et al., 2014), to determine the significance of the direct effect (ADE), indirect effect (AME), and total effect. In each case, we conducted mediation analyses separately for Familiarity/Contact with Transgender People and Openness Towards Non-Binary Gender.

Mediation results were mixed. For the Who-Said-What task,

[8] As we had already presented all possible participant group comparisons in ratings of within-identity pairs, we do not describe them here.

[9] In our Stage 1 Registered Report, we incorrectly said we would include two three-way interactions in these models. There is only one possible three-way interaction, which we include.

Table 14
Mediation Analyses.

	Task	Description	Step 1: Level-2 Mediation DV	Step 2: Define Relevant Group Comparison	Step 3: Mediation Analyses	
					Familiarity with Transgender People	Openness to Non-Binary Gender
(1)	Who-Said-What	Transgender adults made significantly fewer errors than cisgender adults	Total number of errors	Cisgender adults (reference group) vs. Transgender adults	No Mediation AME = −0.05, p = .713 ADE = −0.82, p = .013 Total Effect = −0.87, p = .004	Full Mediation AME = −0.34, p = .045 ADE = −0.54, p = .122 Total Effect = −0.87, p = .004
(2)	Similarity	Cisgender adults rated groups less similar overall than transgender adults, transgender youth, and cisgender youth	Mean similarity rating	Cisgender adults (reference group) vs. All other participants	Partial Mediation AME = 1.29, p < .001 ADE = 8.01, p < .001 Total Effect = 9.30, p < .001	Partial Mediation AME = 3.58, p < .001 ADE = 5.72, p < .001 Total Effect = 9.31, p < .001
(3)	Who-Said-What	Adults had a greater tendency than youth to make within-identity (rather than between-identity) errors	Number of within-identity errors – number of between-identity errors	Adults (reference group) vs. Youth	No Mediation AME = 0.03, p = .392 ADE = −0.58, p = .022 Total Effect = −0.56, p = .028	No Mediation AME = 0.09, p = .200 ADE = −0.65, p = .014 Total Effect = −0.56, p = .029
(4a)	Similarity	Cisgender adults rated within-identity pairs as less similar than cisgender youth	Mean rating of within-identity pairs	Cisgender adults (reference group) vs. Cisgender youth	Partial Mediation AME = 5.67, p < .001 ADE = 11.37, p < .001 Total Effect = 17.03, p < .001	Partial Mediation AME = 12.46, p < .001 ADE = 4.56, p = .041 Total Effect = 17.02, p < .001
(4b)	Similarity	Cisgender youth rated within-identity pairs as less similar than transgender youth and transgender adults	Mean rating of within-identity pairs	Cisgender youth (reference group) vs. Transgender participants	Partial Mediation AME = 2.15, p < .001 ADE = 7.87, p < .001 Total Effect = 10.02, p < .001	Partial Mediation AME = 2.94, p < .001 ADE = 7.07, p < .001 Total Effect = 10.02, p < .001
(5)	Similarity	Cisgender adults rated within-modality pairs as less similar than transgender adults, transgender youth, and cisgender youth	Mean rating of within-modality pairs	Cisgender adults (reference group) vs. All other participants	No Mediation AME = −0.25, p = .637 ADE = 8.36, p < .001 Total Effect = 8.11, p < .001	No Mediation AME = 1.18, p = .182 ADE = 6.94, p = .001 Total Effect = 8.12, p < .001
(6)	Similarity	Transgender youth rated between-both pairs as less similar than cisgender youth and cisgender adults	Mean rating of between-both pairs	Transgender youth (reference group) vs. Cisgender participants	Partial Mediation AME = 1.51, p < .001 ADE = 6.87, p < .001 Total Effect = 8.38, p < .001	Partial Mediation AME = 1.96, p = .002 ADE = 6.41, p = .001 Total Effect = 8.37, p < .001

Note: Rather than including mediation diagrams, we here include the relevant coefficients from such diagrams. AME represents the average mediation effect, ADE represents the average direct effect, and TE represents the total effect.

Familiarity/Contact with Transgender People did not mediate either participant group effect. Openness Towards Non-Binary Gender fully mediated the pattern of transgender adults making fewer errors than cisgender adults, but did not mediate the pattern of adults encoding gender identity more than youth. In short, we saw evidence for mediation in only one of the four cases of the Who-Said-What task.

For Similarity Judgments, both Familiarity/Contact with Transgender People and Openness Towards Non-Binary Gender partially

mediated four of the five effects (or 8 of 10 total effects). However, neither individual difference measure explained the tendency of cisgender adults to rate within-modality pairs as less similar than the other three groups.

4.3. Discussion

Overall, Study 2's results are consistent with our preregistered predictions and demonstrate evidence of some group differences in categorization. We replicated Study 1's general patterns of results among a new sample of cisgender adults, including the novel finding that participants' implicit categorization was characterized by both within-identity *and* within-modality errors (see Supplementary Material). Though effects of gender modality tended to be smaller than effects of gender identity, this result is the first we are aware of to show that people automatically encode both gender identity and gender modality. In all cases, the most common errors observed on the Who-Said-What task were those in which people shared a gender identity and a gender modality (i.e., confusing a transgender woman for another transgender woman), again demonstrating that both aspects of gender were encoded by participants.

The general pattern of Who-Said-What task performance among cisgender adults generalized to our other three participant groups. All groups showed the same general pattern wherein they most often made within-gender and within-modality errors. Despite this commonality, there were a few small group differences. Specifically, the tendency to make more within-identity than between-identity errors was stronger among adults than youth. One interpretation of this group-based differences is that automatic categorization becomes stronger with practice, as adults would have had more experience than youth in categorizing people by gender in daily life (Fiske, 1998); another possibility is that youth rely less on physiological cues to gender than adults (i.e., the faces themselves). Another possibility is that there's an emerging cohort effect wherein young people are generally encoding gender less than mature adults.

In addition, transgender adults were more accurate overall than cisgender adults, while youth groups did not differ significantly from each other or either adult group. As this group-based difference was fully mediated by participants' Openness Towards Non-Binary Gender (not Familiarity/Contact with Transgender People), one interpretation of these results is that transgender adults tend to view gender as a more individual feature, attending to individual-level features rather than groups/labels as part of a recognition that gender can be fluid and flexible (see Atwood et al., 2024; Parker & Igielnik, 2020). More work should follow up on this finding, but the current results suggest there is much-needed nuance to theories about how people automatically categorize the gender of other people they meet in daily life, and how people's attitudes and experiences shape these processes.

Similarity Judgments also replicated the same general pattern from Study 1, although these responses showed more variation between participant groups. Cisgender participants (especially adults) rated within-identity pairs (e.g., transgender and cisgender women) as less similar than did transgender participants, and this was partially mediated by both Familiarity/Contact with Transgender People and Openness to Non-Binary Gender. This result is consistent with the possibility that becoming familiar with trans people makes one able to see that trans people are similar to cis people who share their gender identity (or perhaps that those who see cis and trans people of the same gender as more similar are more comfortable spending time with trans people; other explanations may also apply). Similar explanations may be at play for those who have greater Openness Towards Non-Binary Gender.

Transgender youth rated between-both pairs (e.g., transgender women and cisgender men) as less similar than cisgender participants, and this was partially mediated by both Familiarity/Contact with Transgender People and Openness Towards Non-Binary Gender. Again, perhaps familiarity or openness alerts some individuals to differences between those who share neither a gender identity or gender modality. Finally, cisgender adults rated within-modality pairs (e.g., transgender women and men) as less similar than the other groups did, but this was not mediated by familiarity or attitudes. These within-modality judgments might vary by some other factor we did not measure; for example, people with more awareness of shared experiences of discrimination might judge transgender men and women as more similar. Future research might examine which other aspects of experience predict this type of explicit categorization.

Despite the role of attitudes and experiences in shaping both implicit and explicit categorization, responses on these tasks were largely uncorrelated, as in Study 1 and in line with previous evidence of dissociations between implicit and explicit measures in the gender domain (Devine, 2001; Eidson & Coley, 2014; Greenwald & Banaji, 1995; Kahneman, 2003; Moss-Racusin et al., 2012; Sloman, 1996). In one notable exception, cisgender participants who made more within-modality errors on the Who-Said-What task (e.g., attributing a statement by a transgender man to a transgender woman, or attributing a statement by a cisgender woman to a cisgender man) also judged within-modality pairs (i.e., transgender men and women, cisgender men and women) as more similar. One possibility is that this group of participants might vary more in overall awareness of or focus on gender modality than the other groups, which could affect both their implicit and explicit categorization; future research might explore this possibility.

Overall, the four participant groups differed as expected on both measures of attitudes and experiences that we expected might mediate categorization judgments, with transgender participants and younger participants scoring higher on both Familiarity/Contact with Transgender People and Openness to Non-Binary Gender. Both of these factors statistically mediated some of the group differences observed (in 9 of 14 cases we found evidence for full or partial statistical mediation by these factors), more often statistically mediating effects on the more explicit measure than the more implicit one. These results suggest that greater Familiarity/Contact with Transgender People and greater Openness Towards Non-Binary Gender may help explain why we observed both age and modality-related differences on our categorization tasks. However, because these were cross-sectional analyses, causal conclusions cannot be drawn. Instead, these results provide some preliminary suggestion of important future directions for understanding the kinds of experiences that might lead to different patterns of categorization.

5. General discussion

The goal of the current studies was to shed light on how people process gender-relevant categories in daily life, and to ask if life experiences might influence those processes. By investigating implicit and explicit gender-relevant categorization along dimensions of both gender identity (men and women) and gender modality (transgender and cisgender), in samples of transgender and cisgender youth and mature adults, this work helps refine our psychological theories of gender categorization so that they more accurately reflect the landscape of gender categories in modern society. Specifically, although we observed similar patterns in implicit and explicit gender-relevant categorization across all groups, the existence of group-based differences highlights the importance of experience for gender categorization. This finding is particularly notable because some past theorizing has argued that gender categorization is especially inevitable and more difficult to influence, compared with other types of social categorization (e.g., racial categorization, coalitional categorization; Kurzban et al., 2001). Rather, the current results lend further support to the proposal that individual experience with categories—including gender—can impact people's concepts of categories and category members (Bigler & Liben, 2007; Malpass & Kravitz, 1969; Pollak & Kistler, 2002; Wright & Sladden, 2003; Wright & Stroud, 2002; Zhou et al., 2019). Here we found that much of the variation in gender categorization was partially or fully mediated by variation in familiarity with transgender people and views

of gender fluidity, across participant groups differing in both gender modality and generational age group (Jones, 2021; Minkin & Brown, 2021).

Younger participants and transgender participants saw shared-identity gender groups (e.g., transgender and cisgender women) as more similar on the explicit measure, which was partially mediated by both familiarity with transgender people and beliefs in gender fluidity. This suggests that these experiences are associated with an emphasis on gender identity over and above other aspects of gender (e.g., modality), rather than a reduced focus on gender categorization overall (this could be represented either in lower similarity ratings across the board, or in similarity ratings that did not differ by which pair was being rated).

On the implicit measure, we saw that adult participants had a greater tendency to rely on target gender identity than did youth participants – a pattern that was not mediated by familiarity or openness to non-binary gender. We think this apparent disjunction between the implicit and explicit results may be explained by the differences in the tasks – the Who-Said-What task represented target gender identity both in a label and in a stereotypically gendered face, while the Similarity Judgments used only a label. It may be that adults, compared to youth, use the face-cue to a greater extent, either because they have more years of experience detecting gender from faces or because they consider someone's face a more reliable cue to that person's gender (youth may, for example, know more people whose gender identity does not stereotypically 'match' their appearance). To our knowledge this is the first indication of an emerging possible generational difference in gender categorization. This is consistent with other work (e.g., polling) that suggests younger people, even cisgender younger people, may be thinking about gender in more expansive ways than prior generations (Parker & Igielnik, 2020).

The current work has several limitations. For example, we chose to use color-coded labels to mark gender modality because we knew that gender identity was salient through other physical cues (e.g., physical appearance, hair styles), and so we wanted to ensure that the gender modality dimension was also highlighted in a salient way. Labeling faces as cisgender or transgender is a method used in previous work to capture people's pre-existing ideas about transgender and cisgender people while controlling for visual cues in the actual stimuli (e.g., Howansky et al., 2020; Mao et al., 2019; Wittlin et al., 2018). Nonetheless, this approach of labeling lacks face validity as we do not typically walk around the world with words orienting us to people's gender modality.

The stimuli themselves were also fairly stereotypic portrayals of gender, due to similar concerns about categorization of less stereotypical faces (Gerhardstein & Anderson, 2010; Stern & Rule, 2018; Strauss et al., 2012). Further, our stimuli included only binary gender categories – cisgender and transgender women and men – omitting groups of people who do not identify inside one of these categories (e.g., non-binary people, demi-girls, etc). Future research might test the boundaries of the current findings by using a wider range of faces, visual cues, and gender categories.

Additionally, we used different recruitment strategies for participants across different groups, especially between our youth and adult samples. This conflation of age and recruitment was a necessary part of our initial design, given the difficulty of obtaining such a large sample of transgender youth. However, we recruited two additional samples via Prolific to ensure that recruitment venue alone could not explain group differences in task performance, and we report additional analyses to address this in the Supplementary Material. Our dominant conclusion is that recruitment approach alone does not explain the observed effects. To reduce other differences that would co-occur with recruitment site, we took other precautions - for example, we systematically recruited adults through Prolific in monthly batches that were equal in size to our longitudinal youth sample, to try to match their exposure to recent national and world events that impact treatment. This was necessary given rapid and salient changes in the social and political discourse about transgender rights in the U.S.

Finally, we defined "transgender" participants in a broad way in the current work, even though not all participants in our transgender samples necessarily use this label to describe themselves. We group these participants together here to capture the experiences they share (and do not share with the cisgender samples) of identifying with and/or enacting a gender category that does not align with the one they were assigned at birth. This approach has been utilized in other recent work (e.g., Atwood et al., 2024). However, we acknowledge that this choice cannot capture some important aspects of participants' gender-relevant experiences.

In conclusion, the current findings highlight both commonality and difference in gender categorization as a function of people's own experiences. We further found that people's attitudes about nonbinary identities and experiences with trans people relate to how they categorize other people. Given the central role that gender categories play in shaping how we view ourselves and others, future research will be needed to build on these findings, across a broader range of individuals and experiences, to more closely align our theories of gender categorization with the increasingly varied and nuanced landscape of gender in modern society.

Open practices

All materials and data for Study 1 can be found here: osf.io/pxfus/. All materials and data for Study 2 can be found here: osf.io/qb6mj/, and the Study 2 preregistration can be found here: osf.io/yv9gk.

Author note

Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute Of Child Health & Human Development of the National Institutes of Health under Award Number R01HD092347. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Additional contributions

We thank Megan J. Heise (Princeton University) for reviewing the analytic code for this project.

CRediT authorship contribution statement

Natalie M. Gallagher: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Emily Foster-Hanson:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Kristina R. Olson:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesp.2024.104691.

References

Ashley, F. (2022). "Trans" is my gender modality: A modest terminological proposal. In *Trans Bodies, Trans Selves* (2nd ed.). Oxford University Press.

Atwood, S., Gibson, D. J., Briones Ramírez, S., & Olson, K. R. (2024). Flexibility in continuous judgments of gender/sex and race. *Journal of Experimental Psychology. General*. https://doi.org/10.1037/xge0001512

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using **lme4**. *Journal of Statistical Software, 67*(1).

Bian, L., Leslie, S.-J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science, 355*(6323), 389–391. https://doi.org/10.1126/science.aah6524

Bigler, R. S., & Liben, L. S. (2007). Developmental intergroup theory: Explaining and reducing children's social stereotyping and prejudice. *Current Directions in Psychological Science, 16*(3), 162–166.

Boyer, C. R., & Galupo, M. P. (2018). Transgender friendship profiles: Patterns across gender identity and LGBT affiliation. *Gender Issues, 35*(3), 236–253. https://doi.org/10.1007/s12147-017-9199-4

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

Castelli, L., Tumino, M., & Carraro, L. (2022). Face mask use as a categorical dimension in social perception. *Scientific Reports, 12*(1), 17860. https://doi.org/10.1038/s41598-022-22772-2

Cohen, L. B., & Strauss, M. S. (1979). Concept acquisition in the human infant. *Child Development*, 419–424.

Conley, M. I., Dellarco, D. V., Rubien-Thomas, E., Cohen, A. O., Cervera, A., Tottenham, N., & Casey, B. (2018). The racially diverse affective expression (RADIATE) face stimulus set. *Psychiatry Research, 270*, 1059–1067. https://doi.org/10.1016/j.psychres.2018.04.066

deMayo, B., Kahn-Samuelson, S., & Olson, K. (2021). Endorsement of gender stereotypes in gender diverse and cisgender adolescents and their parents. *PLoS One*. https://doi.org/10.31234/osf.io/rh6mn

Devine, P. G. (2001). Implicit prejudice and stereotyping: How automatic are they? Introduction to the special section. *Journal of Personality and Social Psychology, 81*(5), 757–759. https://doi.org/10.1037/0022-3514.81.5.757

Durwood, L., Gallagher, N., Sifre, R., & Olson, K. R. (2024). A study of parent-reported internalizing symptoms in transgender youth before and after childhood social transitions. *Clinical Psychological Science, 12*(5), 984–996.

Eidson, R. C., & Coley, J. D. (2014). Not so fast: Reassessing gender essentialism in young adults. *Journal of Cognition and Development, 15*(2), 382–392.

Fast, A. A., & Olson, K. R. (2018). Gender development in transgender preschool children. *Child Development, 89*(2), 620–637. https://doi.org/10.1111/cdev.12758

Fine, R. D., Gelman, S. A., & Ho, A. K. (2023). Changing beliefs about gender: The relation between contact with gender nonconforming individuals and gender essentialism. *Psychology & Sexuality, 1–17*. https://doi.org/10.1080/19419899.2023.2181706

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 357–411). McGraw-Hill.

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*, Article 104887. https://doi.org/10.1016/j.cognition.2021.104887

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

Glazier, Jessica J., Gülgöz, Selin, & Olson, Kristina R. (2020). Gender Encoding in Gender Diverse and Gender Conforming Children. *Child Development*. https://doi.org/10.1111/cdev.13399

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*(1), 4–27.

Gülgöz, S., Alonso, D. J., Olson, K. R., & Gelman, S. A. (2021). Transgender and cisgender children's essentialist beliefs about sex and gender identity. *Developmental Science, 24*(6). https://doi.org/10.1111/desc.13115

Gülgöz, S., DeMeules, M., Gelman, S. A., & Olson, K. R. (2019). Gender essentialism in transgender and cisgender children. *PLoS One, 14*(11), Article e0224321. https://doi.org/10.1371/journal.pone.0224321

Gülgöz, S., Glazier, J. J., Enright, E. A., Alonso, D. J., Durwood, L. J., Fast, A. A., … Olson, K. R. (2019). Similarity in transgender and cisgender children's gender development. *Proceedings of the National Academy of Sciences, 116*(49), 24480–24485. https://doi.org/10.1073/pnas.1909367116

Gülgöz, S., Glazier, J. J., Enright, E. A., Alonso, D. J., Durwood, L. J., Fast, A. A., … Olson, K. R. (2019). Similarity in transgender and cisgender children's gender development. *PNAS*. https://doi.org/10.1073/pnas.1909367116

Herman, J. L., Flores, A. R., & O'Neill, K. K. (2022). *How many adults and youth identify as transgender in the United States?*.

Howansky, K., Albuja, A., & Cole, S. (2020). Seeing gender: Perceptual representations of transgender individuals. *Social Psychological and Personality Science, 11*(4), 474–482. https://doi.org/10.1177/1948550619875143

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

Jones, J. M. (2021). *LGBT identification rises to 5.6% in latest U.S. Estimate*. Gallup. https://news.gallup.com/poll/329708/lgbt-identification-rises-latest-estimate.aspx.

Judd, C. M., McLelland, G. H., & Ryan, C. S. (2017). *Data analysis: A model comparison approach to regression, ANOVA, and beyond* (3rd ed.). Routledge. https://doi.org/10.4324/9781315744131

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*(9), 697–720. https://doi.org/10.1037/0003-066X.58.9.697

Kassambara, A. (2022). *rstatix: Pipe-friendly framework for basic statistical tests* [computer software]. https://CRAN.R-project.org/package=rstatix.

Klauer, K. C., & Wegener, I. (1998). Unraveling social categorization in the "who said what?" paradigm. *Journal of Personality and Social Psychology, 75*(5), 1155–1178.

Kteily, N. S., Hodson, G., Dhont, K., & Ho, A. K. (2019). Predisposed to prejudice but responsive to intergroup contact? Testing the unique benefits of intergroup contact across different types of individual differences. *Group Processes & Intergroup Relations, 22*(1). https://doi.org/10.1177/1368430217716750

Kung, K. T. F. (2021). Preschool gender-typed play behavior predicts adolescent gender-typed occupational interests: A 10-year longitudinal study. *Archives of Sexual Behavior, 50*(3), 843–851. https://doi.org/10.1007/s10508-021-01976-z

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences, 98*(26), 15387–15392. https://doi.org/10.1073/pnas.251541498

Leinbach, M. D., & Fagot, B. I. (1993). Categorical habituation to male and female faces: Gender schematic processing in infancy. *Infant Behavior and Development, 16*(3), 317–332. https://doi.org/10.1016/0163-6383(93)80038-A

Lenth, R. V. (2021). emmeans: Estimated marginal means, aka least-squares means (version R package version 1.6.2-1) [computer software]. https://CRAN.R-project.org/package=emmeans.

Levesque, H., Davis, E., & Morgenstern, L. (2012). The Winograd Schema challenge. In *Proceedings of the thirteenth international conference on principles of knowledge representation and reasoning*. Association for the Advancement of Artificial Intelligence.

Lüdecke, D., Bartel, A., Schwemmer, C., Powell, C., Djalovski, A., & Titz, J. (2022). sjPlot: Data visualization for statistics in social science (version R package version 2.8.12) [Computer software]. https://CRAN.R-project.org/package=sjPlot.

Maccoby, E. E. (1988). Gender as a social category. *Developmental Psychology*. https://doi.org/10.1037/11347-009

Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology, 13*(4), 330–334. https://doi.org/10.1037/h0028434

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, C. L., & Ruble, D. (2004). Children's search for gender cues: Cognitive perspectives on gender development. *Current Directions in Psychological Science, 13*(2), 67–70. https://doi.org/10.1111/j.0963-7214.2004.00276.x

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather Homophily in social networks. *Annual Review of Sociology, 27*, 415–444.

Mervis, C. B., & Pani, J. R. (1980). Acquisition of basic object categories. *Cognitive Psychology, 12*(4), 496–522. https://doi.org/10.1016/0010-0285(80)90018-3

Minkin, R., & Brown, A. (2021). *Rising shares of U.S. adults know someone who is transgender or goes by gender-neutral pronouns*. Pew Research Center. https://www.pewresearch.org/fact-tank/2021/07/27/rising-shares-of-u-s-adults-know-someone-who-is-transgender-or-goes-by-gender-neutral-pronouns/.

Molin, A., Simond, A., Sato, S., Jaeggi, T., Gygax, P. M., & Meuwly, N. (2021). 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, 30*(8), 901–914. https://doi.org/10.1080/09589236.2020.1844644

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences, 109*(41), 16474–16479. https://doi.org/10.1073/pnas.1211286109

Muradoglu, M., Cimpian, J. R., & Cimpian, A. (2023). Mixed-effects models for cognitive development researchers. *Journal of Cognition and Development, 24*(3), 307–340. https://doi.org/10.1080/15248372.2023.2176856

Murphy, G. L. (2002). *The big book of concepts*. MIT Press.

Olson, K. R., & Enright, E. A. (2018). Do transgender children (gender) stereotype less than their peers and siblings? *Developmental Science, 21*(4), Article e12606. https://doi.org/10.1111/desc.12606

Olson, K. R., & Gülgöz, S. (2018). Early findings from the TransYouth Project: Gender Development in Transgender Children. *Child Developmental Perspectives, 12*(2), 93–97.

Olson, K. R., Key, A. C., & Eaton, N. R. (2015). Gender cognition in transgender children. *Psychological Science, 26*(4), 467–474. https://doi.org/10.1177/0956797614568156

Palan, S., & Schitter, C. (2018). Prolific.Ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance, 17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Parker, K., Horowitz, J. M., & Brown, A. (2022). *Americans' complex views on gender identity and transgender issues*. Pew Research Center. https://www.pewresearch.org/social-trends/2022/06/28/americans-complex-views-on-gender-identity-and-transgender-issues/.

Parker, K., & Igielnik, R. (2020). *On the cusp of adulthood and facing an uncertain future: What we know about gen Z so far*. Pew Research Center. https://www.pewresearch.org/social-trends/2020/05/14/on-the-cusp-of-adulthood-and-facing-an-uncertain-future-what-we-know-about-gen-z-so-far-2/.

Pollak, S. D., & Kistler, D. J. (2002). Early experience is associated with the development of categorical representations for facial expressions of emotion. *Proceedings of the National Academy of Sciences, 99*(13), 9072–9076. https://doi.org/10.1073/pnas.142165999

Quinn, P. C., Yahr, J., Kuhn, A., Slater, A. M., & Pascalis, O. (2002). Representation of the gender of human faces by infants: A preference for female. *Perception, 31*(9), 1109–1121. https://doi.org/10.1068/p3331

Revelle, W. (2019). *Psych: Procedures for psychological, psychometric, and personality research*. [Computer software]. Northwestern University.
Ridout, N., Astell, Reid, I., Glen, T., & O'Carroll, R. (2003). Memory bias for emotional facial expressions in major depression. *Cognition & Emotion, 17*(1), 101–122.
Rosch, E. (1978). Principles of categorization. In E. Rosch, & B. B. Lloyd (Eds.), *Cognition and categorization*. Lawrence Erlbaum Associates, Inc.
Rubin, J. D., Atwood, S., & Olson, K. R. (2020). Studying gender diversity. *Trends in Cognitive Sciences, 24*(3), 163–165. https://doi.org/10.1016/j.tics.2019.12.011
Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (2006). Gender development. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (pp. 858–932). John Wiley & Sons, Inc.
Serano, J. (2007). *Whipping girl (second edition)*. Seal Press. https://www.hachettebookgroup.com/titles/julia-serano/whipping-girl/9781580056229/?lens=seal-press.
Sesko, A. K., & Biernat, M. (2010). Prototypes of race and gender: The invisibility of Black women. *Journal of Experimental Social Psychology, 46*(2), 356–360. https://doi.org/10.1016/j.jesp.2009.10.016
Shutts, K., Kenward, B., Falk, H., Ivegran, A., & Fawcett, C. (2017). Early preschool environments and gender: Effects of gender pedagogy in Sweden. *Journal of Experimental Child Psychology, 162*, 1–17. https://doi.org/10.1016/j.jecp.2017.04.014
Shutts, K., Roben, C. K. P., & Spelke, E. S. (2013). Children's use of social categories in thinking about people and social relationships. *Journal of Cognition and Development, 14*(1), 35–62. https://doi.org/10.1080/15248372.2011.638686
Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3–22.
Stangor, C., Lynch, L., Duan, C., & Glass, B. (1992). Categorization of individuals on the basis of multiple social features. *Journal of Personality and Social Psychology, 62*(2), 207–218.
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
Strauss, M. S., Newell, L. C., Best, C. A., Hannigen, S. F., Gastgeb, H. Z., & Giovannelli, J. L. (2012). The development of facial gender categorization in individuals with and without autism: The impact of typicality. *Journal of Autism and Developmental Disorders, 42*(9), 1847–1855. https://doi.org/10.1007/s10803-011-1428-1
Taylor, M. G. (1996). The development of children's beliefs about social and biological aspects of gender differences. *Child Development, 18*.
Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology, 36*(7), 78–793.
Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). **Mediation**: *R* package for causal mediation analysis. *Journal of Statistical Software, 59*(5). https://doi.org/10.18637/jss.v059.i05
Voci, A., & Hewstone, M. (2003). Intergroup contact and prejudice toward immigrants in Italy: The mediational role of anxiety and the moderational role of group salience. *Group Processes & Intergroup Relations, 6*(1), 37–54. https://doi.org/10.1177/1368430203006001011
Weeks, M., & Lupfer, M. B. (2004). Complicating race: The relationship between prejudice, race, and social class categorizations. *Personality and Social Psychology Bulletin, 30*(8), 972–984. https://doi.org/10.1177/0146167204264751
Weinraub, M., Clemens, L. P., Sockloff, A., Ethridge, T., Gracely, E., & Myers, B. (1984). The development of sex role stereotypes in the third year: Relationships to gender labeling, gender identity, sex-types toy preference, and family characteristics. *Child Development, 55*(4), 1493. https://doi.org/10.2307/1130019
Weisman, K., Johnson, M. V., & Shutts, K. (2015). Young children's automatic encoding of social categories. *Developmental Science, 18*(6), 1036–1043. https://doi.org/10.1111/desc.12269
Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43).
Wittlin, N. M., Dovidio, J. F., LaFrance, M., & Burke, S. E. (2018). About face: Memory for transgender versus cisgender targets' facial appearance. *Journal of Experimental Social Psychology, 78*, 77–92. https://doi.org/10.1016/j.jesp.2018.04.009
Wright, D. B., & Sladden, B. (2003). An own gender bias and the importance of hair in face recognition. *Acta Psychologica, 114*(1), 101–114. https://doi.org/10.1016/S0001-6918(03)00052-0
Wright, D. B., & Stroud, J. N. (2002). Age differences in lineup identification accuracy: People are better with their own age. *Law and Human Behavior, 26*(6), 641–654. https://doi.org/10.1023/A:1020981501383
Xie, Y. (2023). *Knitr: A general-purpose package for dynamic report generation in R* (version R package version 1.42) [computer software]. https://yihui.org/knitr/.
Younger, B. A., & Fearing, D. D. (1999). Parsing items into separate categories: Developmental change in infant categorization. *Child Development, 70*(2), 291–303. https://doi.org/10.1111/1467-8624.00022
Zhou, X., Elshiekh, A., & Moulson, M. C. (2019). Lifetime perceptual experience shapes face memory for own- and other-race faces. *Visual Cognition, 27*(9–10), 687–700. https://doi.org/10.1080/13506285.2019.1638478
Zhu, H. (2024). kableExtra: Construct complex table with "kable" and pipe syntax [computer software]. https://CRAN.R-project.org/package=kableExtra.
Zosuls, K. M., Ruble, D. N., Tamis-LeMonda, C. S., Shrout, P. E., Bornstein, M. H., & Greulich, F. K. (2009). The acquisition of gender labels in infancy: Implications for gender-typed play. *Developmental Psychology, 45*(3), 688–701. https://doi.org/10.1037/a0014053