LEARNING TO COMPREHEND AND PRODUCE SINGULAR *THEY*

By

Bethany Gardner

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

August 11, 2023

Nashville, Tennessee

Approved:

Sarah Brown-Schmidt, Ph.D.

Melissa Duff, Ph.D.

Lisa Fazio, Ph.D.

Duane Watson, Ph.D.

ACKNOWLEDGEMENTS

I gratefully acknowledge NSF Linguistics (Doctoral Dissertation Research Improvement #2214299), as well as the small grants programs of the Vanderbilt LGBT Policy Lab and the Peabody Office of Equity, Diversity, and Inclusion for their support of the experiments presented here. The views presented in this dissertation are my own, and do not necessarily reflect the views of these organizations.

Thank you also to Morrigan Dunlap-Loomis, who contributed to annotating data for Experiment 3 and collecting data for Experiment 4.

I first have to thank my advisor, Sarah Brown-Schmidt. In one of our first conversations about what would eventually become this project, she exclaimed "I *love* pronouns! What are the kids these days doing with pronouns?" Since then, she has continued to be both excited about my research questions and pushing me to make them better. The structure of grad school can make it feel like you're never doing enough, and the thing I have appreciated most about her as a mentor is that she assumes you're doing the best you can, instead of expecting people to continuously prove that they're trying hard enough.

At Vanderbilt, I also have to acknowledge Joint Lab Meeting, the group of cognitive psych people who heard about these experiments starting from the project proposal stage all the way to the practice talk stage, and asked me a lot of questions that helped clarify my thinking and explanations. You all are a big reason this exists in a form that makes sense outside of my brain, thanks! Thank you also to the grad student writing group that has formed over the past year and a half, especially but not exclusively: Melissa, Raunak, Rebecca, and Camille. Having other people to write with kept my brain from convincing me that everyone else was doing fine and I was the only one finding writing hard, slow work. I'm also grateful for McKay's camaraderie: having multiple breakdowns trying to do stats homework in SPSS-via-VMWare on a zoom call truly was one of the peak PhD student experiences.

I owe a lot of thanks to Chigusa Kurumada, my undergrad advisor at Rochester, who was a dedicated mentor and a consistent advocate for my success. Looking back, she treated several of us senior undergrads like first-year grad students, which gave me so many learning opportunities early on and helped me figure out that I really do love psycholinguistics research.

Although this dissertation work is all quantitative, learning about sociolinguistics from the queer linguists I met at the LSA summer school and through linguistics Twitter has enriched my thinking tremendously. I love this community's sense of delight about language and everything new that people do with it, which has kept me excited about research. Kirby Conrod's essay "the president of pronouns gives you permission to just kind of try stuff out" helped me decide to explore they/them as a researcher and as a person.

Luis and Ariel, thanks for convincing me that I am both more underpaid and more qualified than I thought. Academia can be such a weird bubble of expectations sometimes, and you two especially were my voices of reason from the outside.

I'm also grateful to my dad, who was one of the only adults in my life as a teenager who treated my constant questions and drive to learn new things as a gift, not as unfeminine or a threat to authority. I didn't end up where he expected, but I wouldn't have ended up here without him either.

There are a lot of other people who are less directly connected to my education or career, but without whom I wouldn't have made it this far. The biggest one is my medical team. I developed what I now know is post-viral dysautonomia my senior year of undergrad, then started grad school in the "sounds like you're just an anxious woman, have you tried yoga?" stage of chronic illness diagnosis. To the doctors who actually listened to me, then kept trying until we found something that worked: thank you. Special thanks to the GI surgeon who did the procedure that basically gave me my life back, but who was really not expecting a patient to reference the paper she published about the success rates of the surgery. I also owe a lot of thanks to the Philadelphia Federation of Teachers, my dad's union. I was able to keep their health insurance through the majority of grad school, and their advocacy meant that I was able to afford the healthcare I needed. I'm also grateful that my advisor and my department were understanding and helpful about disability accommodations, especially at the times I needed them most, which was before I escaped from diagnostic limbo and could get the official documentation to show to Vanderbilt.

Ariel, thanks for being the best friend I could ask for and always one text away. She has been joking that she should get a PhD for supporting her partner and her best friend through PhDs at the same time. In lieu of honorarily granting her a third of this degree, this dissertation is dedicated to her.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Introduction: The State of Singular *They*

0.1 Overview

As the use of they/them pronouns becomes more common and more accepted, an increasing number of speakers find themselves making seemingly paradoxical errors, such as in the following headline:

> *[1] Demi Lovato$_i$ thanks Lizzo for correcting paparazzo who misgendered her$_i$: "I love you."*
> *The "Dancing With The Devil" singer$_i$ came out as nonbinary in May and goes by "they/them" pronouns.*
> — *Entertainment Weekly* (Towers, 2021)

The article describes a press interview where Lizzo corrects paparazzi misgendering Lovato. In the quoted exchange, Lizzo is asked a question about collaborating with Lovato that uses she/her to refer to Lovato. Lizzo corrects this, the question is repeated still using she/her, and Lizzo states twice more that Lovato uses they/them. Although the paparazzi did not change their pronoun use in this conversation, Lovato was grateful for the support. Despite the article being about misgendering and they/them pronouns, the author does not consistently use they/them to refer to Lovato. However, it is unlikely that the author entirely refuses to use singular *they* or the publication prohibits it, as several references to Lovato in the middle of the article are correct. Instead, the author appears to not have noticed the alternations between she/her and they/them, despite the context making accurate use of they/them pronouns as salient as possible, and despite good intentions.

One of the starting questions of this dissertation is *why do people do this?* Why are errors like "she uses they/them pronouns" common, where the speaker knows that a particular person uses they/them, recalls it in the conversational context, and expresses support for that person's identity, but still produces *she* instead of *they* when referring to them? The growing support for and familiarity with singular *they* during the past few years suggests that many people are open

to learning (Minkin & Brown, 2021; Parker et al., 2019), and that a substantial part of the barrier causing frequent misgendering comes from lower-level language processes for producing pronouns, as opposed to higher-level attitudes about gender and language.

In addition to learning that the grammatical representation of *they* allows it to corefer with specific singular referents (i.e., proper names), the learning process may require a change from accessing a person's pronoun (*he* or *she*) based on semantic/conceptual features of a person (e.g., Antón-Méndez, 2010), or based on morphosyntactic gender marking associated with a person's name (e.g., Schmitt et al., 1999). The first three experiments argue for a model where, in order to use singular *they* correctly, speakers may instead need to retrieve information from episodic memory about a person's stated pronouns or which pronouns other speakers use to refer to them. Experiment 1 establishes a measure for how people learn to associate pronouns with a person, and how they use this information to select which pronouns to produce. Experiment 2 investigates how providing people with information about why paying attention to gendered language is important and how seeing singular *they* modeled can support memory for and production of they/them pronouns. Experiment 3 moves from written to spoken production, testing how including pronouns on nametags and in introductions—common EDI recommendations—affects pronoun choice. Finally, Experiment 4 investigates online comprehension, using the visual world paradigm to characterize the processing of *they* coreferring with proper names. Before that, the remainder of this introduction reviews terminology about coreference and gender; prior research about the acceptability, comprehension, and production of singular *they*; and the sociopolitical context of misgendering.

0.2 Terminology

0.2.1 Coreference

In conversations and writing, we use a variety of expressions to refer to the same entity, and keeping track of this—doing coreference resolution—is one of the core problems of language

comprehension (Cornish, 2006; Garnham & Cowles, 2006; Huang, 2006). When talking about people, speakers have the choice between proper names (*Bethany, Bethany Gardner*), pronouns (*they, she*), role nouns (*the author*), and other noun phrase descriptions (*the author of this dissertation*). Comprehenders have to identify who is being talked about when a person is first mentioned, then match later referring expressions back to that referent. Pronouns nearly always refer to someone mentioned previously—an antecedent—instead of introducing a new referent (Kennison et al., 2009). However, in many contexts there are multiple possible antecedents for the same pronoun, and understanding how comprehenders successfully manage this has been the focus of decades of psycholinguistics research. As a brief summary, one class of theories argues that comprehenders identify the referent based on which entities are currently "centered" in attention, which is determined primarily based on grammatical roles (e.g., the subject of the sentence) and information structure (e.g., topic changes) (Gordon et al., 1993; Grosz & Sidner, 1986; Joshi et al., 2006; Walker, 2002). The second major class of theories, coherence-driven approaches, argues that coreference resolution is a consequence of broader inferential reasoning using world knowledge to construct an interpretation of the conversation as a whole (Greene et al., 1992; McKoon et al., 1996). While centering and coherence approaches have generally been considered mutually exclusive, one way of reconciling them is a probabilistic model where comprehenders are using coherence-driven expectations about what entities the speaker will refer to next, and speakers are using centering-driven likelihoods to choose referential forms (Kehler & Rohde, 2013). Aside from the terminology used for reference resolution, the key takeaway for the rest of this dissertation is that coreference—even for neutral, established, uncontroversial pronouns like *he* and *she*—is complicated.

0.2.2 Gender In Linguistics

In a linguistics context, *gender* can refer to multiple concepts: grammatical, conceptual, and social. First, grammatical gender refers to linguistic features. Languages with grammatical

gender systems (e.g., Spanish, French, and German) classify nouns by gender; and determiners, pronouns, and adjectives have morphosyntactic features that mark agreement with the noun's gender (Gygax et al., 2019). For example, *book* is a masculine noun in Spanish (*libro*$_{MASC}$), so the masculine forms of determiners are used (e.g., *el*$_{MASC}$ instead of *la*$_{FEM}$) and adjectives are conjugated with the masculine ending (e.g., *nuevo*$_{MASC}$ instead of *nueva*$_{FEM}$), resulting in phrases like *el*$_{MASC}$ *libro*$_{MASC}$ *nuevo*$_{MASC}$ for *the new book.* Here, *gender* means something more like *kind* or *class*, and while many grammatical gender languages label the classes as masculine, feminine, and neutral, some languages have more than two or three genders (Corbett, 2013a). Most gender classifications of nouns in grammatical gender languages are clearly arbitrary—there is nothing about a book that makes it masculine or feminine. Instead, grammatical gender can be thought of as a component of the word's structure, not its meaning (Antón-Méndez et al., 2002; Wang & Schiller, 2019).

Languages without grammatical gender, such as English, are typically referred to as natural gender languages. Natural gender languages do not group nouns into gender classes, but many do mark gender for pronouns referring to people (Gygax et al., 2019; Siewierska, 2013). Gender marking of pronouns in natural gender languages is so-called because it is assumed to reflect inherent biological categories (e.g., Corbett, 2013b; critiqued by McConnell-Ginet, 2014). This becomes a problem if you acknowledge that gender (and to some degree, biological sex) is a social construction with the potential to have more than two binary categories. With regard to pronouns, a better explanation—in terms of gender theory as well as simply explaining linguistic behavior—is to treat pronouns as reflecting a shared social understanding, instead of an ontological or biological claim (Conrod, 2019b, discussed in more detail later in §0.5). With regard to nouns, instead of calling this natural gender, I refer to this as conceptual gender, following Ackerman (2019). In addition to accounting for modern concepts of social gender, decoupling gendered language from claims about inherent natural categories also facilitates making distinctions between two different ways that nouns can carry conceptual gender information,

which become relevant in psycholinguistics experiments. In definitionally-gendered nouns, the gender information is part of the meaning itself, e.g., *son* or *queen.* In stereotypically-gendered nouns, gender information is part of our knowledge about gender stereotypes, distributions, and expectations, e.g., *mechanic* or *nurse.* Even if they do not believe this should be the case, people know that being a mechanic fits better into expectations for men and being a nurse fits better into expectations for women, and that people employed as mechanics are more likely to be men and people employed as nurses are more likely to be women. Similarly, English first names carry probabilistic information about gender. When talking about names for the design of the experiments in this dissertation, I refer to people's knowledge about gender associations of a name (i.e., that most people named *Mary* are women, or that people named *Jordan* are commonly men or women).[1]

Part of coreference resolution requires evaluating whether a potential antecedent matches the pronoun for features including gender and number. Which aspect(s) of gender, precisely, are being evaluated? Grammatical, conceptual, or social? Since modern English does not have grammatical gender, Ackerman (2019) argues that we are using conceptual gender (see Sato et al., 2013 for similar empirical findings). In order to distinguish how grammatical, conceptual, and social gender may be cognitively and linguistically encoded, Ackerman presents a three-tier model of gender agreement. The exemplar tier represents knowledge about our experiences with cues about gender, including the gender associations of first names, clothing and hair styles, vocal pitch, and other speech characteristics indexed as masculine or feminine. The distribution on the exemplar tier is strongly bimodal for most people, but would change as someone gains more experience with gender-nonconforming and nonbinary people. The category tier represents our cognitive categorizations about gender, conceptualized as two discrete, non-overlapping

[1] *See Bjorkman (2017) and Konnelly & Cowper (2020) for a discussion of how proper names may have morphosyntactic gender features used to evaluate agreement with pronouns, a question that this dissertation remains agnostic about.*

categories laid over the bimodal distribution of the exemplar tier. Children develop categories for gender early on, and people automatically categorize others as male or female, regardless of whether these inferences are accurate or necessary (Bussey & Bandura, 1999; Fagot & Leinbach, 1993; Friedman, 2014; Martin et al., 2002; Martin & Ruble, 2004; Waxman, 2010). This tendency in language comprehension is discussed later in §0.4.1. Finally, the feature tier represents labels and grammatical features associated with the categories on the category tier. In summary, being precise about what we mean when we say *conceptual gender* requires making distinctions between the knowledge and experiences we use to make gender inferences, the categories we infer from that knowledge and experience, and the labels we assign to those categories.

0.2.3 Forms of Singular *They*

Forms of singular *they* vary in three primary aspects—definiteness, specificity, and gender information—and the usage becomes more contested as referents become more definite, specific, and gender-known (Bjorkman, 2017; Conrod, 2019b; Konnelly & Cowper, 2020). Broadly speaking, definite forms, including proper names and noun phrases with *the*, allow the comprehender to identify a referent; indefinite forms, including noun phrases with *a/an*, do not (Abbott, 2006). Generic forms refer to a group of people, while specific forms refer to an individual (Carlson, 2006; Farkas, 2006; Fodor & Sag, 1982). Definiteness and specificity are distinct features here, since expressions can be specific but indefinite—referring to one referent within a group, but not specifying which one.

The oldest and most common form of singular *they* is the generic indefinite:

[2] *There's not a man$_i$ I meet but doth salute me*
As if I were their$_i$ well-acquainted friend
— A Comedy of Errors (Shakespeare, 1623)

[3] *Everybody$_i$ was punctual, everybody$_i$ in their$_i$ best looks:*
not a tear, and hardly a long face to be seen.
— Emma (Austen, 1815)

This form dates back to late Middle English, with some of the earliest examples appearing in Chaucer's *Canterbury Tales*, and has been in continuous use since then (Balhorn, 2004; Baskervill & Sewell, 1895; OED Online, 2021). In the 18th century, prescriptive grammarians started pushing for the exclusive use of *he* as a generic (Bodine, 1975), which I discuss more in §0.8. Currently, generic indefinite singular *they* is widely accepted (Hekanaho, 2020). The majority of style guides allow it, but still recommend using the phrase *he or she* or rephrasing the sentence to be plural (Robertson, 2021).

Compared to the generic indefinite, specific definite forms of singular *they* show much more variation in acceptability and frequency. It can be used in several gender-unspecified contexts, such as when the speaker does not know the referent's gender. This context implies that the referent's gender is unknown, not necessarily that they are nonbinary or use they/them pronouns (Conrod, 2019b). The other speaker can add more information, either directly as in [4], or indirectly through modeling pronoun use in a subsequent utterance.

[4] A: Oh, okay, okay, um…What—Would you mind telling me what happened? Uh, what they$_i$ did?
B: He$_i$.
A: Did he$_i$ look like he$_i$ hadn't slept in about a week?
— *The Magnus Archives*, pre-scripted podcast (Sims & Newall, 2019)

Typically, if information about the referent's gender becomes available, speakers will switch to using *he* or *she*. The primary variance I observe is whether people hedge an unknown with *they*, or make a potentially incorrect guess with *he* or *she.* This form is also used in contexts where the speaker knows the gender of the referent and the addressee does not, but the speaker does not include that information. This seems to be most common when referring to people by relationships or contexts instead of by name, since the addressee does not know them (e.g., *my friend from undergrad*). [5] shows a conversation where Speaker B interprets Speaker A's use of singular *they* as gender-unspecified. When Speaker A introduces the referent's name, Speaker B uses

this information to switch from gender-unspecified *they* to specifying *he*. Speaker A then has to clarify that they had meant gender-specified *they*, where the referent uses they/them not he/him:

> *[5] A: A friend$_i$ of mine…they've$_i$ sort of dropped off the face of the earth and I'm worried about them$_i$. I wouldn't normally go this crazy about this, but I wanted to make sure they're$_i$ alright. I just have a weird feeling about it all.*
> *B: What's their$_i$ name?*
> *A: Ben$_i$. Ben Bernard$_i$.*
> *B: Put your number in and I'll let you know if I find him$_i$.*
> *A: Them$_i$.*
> *B: Got it, great.*
> — *The College Tapes*, pre-scripted podcast (Snow & Shippen, 2021)

The specific, definite, and <u>gender-specified</u> form of singular *they*—the context described by "using they/them pronouns"—is the newest and most contested form. Although English first names do not carry definitive gender information, most carry strong probabilistic information and are perceived as gendered. Singular *they* coreferring with a proper name is carrying information about the referent's gender, not leaving it unspecified, e.g.:

> *[6] Howell$_i$ is about to graduate from college, and finding the perfect job isn't necessarily their$_i$ top priority. They're$_i$ thinking holistically about community, joy, and fulfillment. They$_i$ are also looking beyond their$_i$ 20s.*
> — "The one-size-fits-all narrative of your 20s needs to change," *The Atlantic* (Stauffer, 2021)

> *[7] Without warning, Francis$_i$ kicks their$_i$ free leg against the table. It does not move…Their$_i$ chest rises and falls rapidly as they$_i$ are lowered into the only seat, the dusty air of the theater scratching their$_i$ throat and drying their$_i$ mouth.*
> — *The Magnus Archives*, pre-scripted podcast (Sims & Newall, 2020)

> *[8] With my work I think you definitely know the male figure$_i$ is isolated, um, feels unaccepted, feels, uh, um, just that they$_i$ have no identity.*
> — "Other Voices: Matthew Neenan," video interview (Neenan, 2020)

Singular *they* coreferring with <u>definitionally-gendered</u> role nouns behaves similarly (see (Bjorkman, 2017; Konnelly & Cowper, 2020). However, these usages are less common, likely

because people who use they/them pronouns often also prefer gender-neutral titles (e.g., *Mx*) and role nouns (e.g., *sibling*) (Cassian, 2021, 2022). Cases where the speaker unambiguously marks gender in one form of reference but not in pronouns do occur, such as [8]. This usage is not as relevant to the current questions, since speakers are rarely asked to produce it, but what discourse factors cause speakers to spontaneously produce it is a question for future research.

0.2.4 LGBTQ Identity

Here, I use trans and gender diverse (TGD) as an umbrella term for anyone whose gender is different than their sex assigned at birth (what was put on their birth certificate) and cisgender as an umbrella term for anyone whose gender matches their sex assigned at birth. People use a variety of terms to describe their identities, which currently include *transgender, nonbinary, genderqueer, genderfluid, gender non-conforming*, and *queer* (Cassian, 2021, 2022; GLAAD, 2020; Harrison et al., 2012; James et al., 2016; Kuper et al., 2012). Currently, nonbinary is the most common label for people whose gender is not entirely within the categories of male or female, some of whom also identify as trans (Cassian, 2021, 2022; GLAAD, 2020; James et al., 2016; Wilson & Meyer, 2021). While many nonbinary people use they/them pronouns, not all do, and not all people who use they/them are nonbinary. Pronouns are one of the cues we use to convey gender, and they strongly correlate with gender identity, but the mapping is not a one-to-one rule. Because of this, I specify “people who use they/them pronouns” instead of “nonbinary people” or “nonbinary pronouns.” Otherwise, when describing prior research findings, I use the authors’ terminology, noting that variations in the language used to recruit and describe participants select for different subsets of the TGD community.

0.3 Acceptability and Grammaticality

Much of the recent linguistics research on singular *they* has focused on acceptability and grammaticality judgments. People’s acceptability judgments tend to fall into one of three patterns:

"Non-innovators" only accept generic antecedents for singular *they* (e.g., *each person, everyone*) and do not accept any specific antecedents. In addition to generic antecedents, "innovators" accept specific, but only gender-unspecified antecedents for singular *they* (e.g., *my friend*). To accommodate proper names, "innovative" speakers can add a new version of the name to their linguistic knowledge that is not grammatically marked as masculine or feminine. This means that they can accept singular *they* for a person that they have learned uses they/them pronouns, but implies that it occurs through an exception to their gender agreement system, not a change to it. Finally, "super-innovators" accept any singular antecedents for *they*, including gender-specified ones (e.g., definitionally-gendered nouns like *sister,* any proper names) (Bjorkman, 2017; Camilliere et al., 2021; Konnelly & Cowper, 2020).

Several factors have been repeatedly demonstrated to correlate with acceptability judgments. People who are younger (Conrod, 2019b; Hekanaho, 2020; Parker et al., 2019), who are LGBTQ+ (Hernandez, 2020; Nichols et al., 2019), and who know trans and gender diverse people (Ackerman et al., 2017; Hekanaho, 2020) consider singular *they* more acceptable. Attitudes about gender play a role: unsurprisingly, people who express more support of TGD people are more likely to consider singular *they* acceptable (Hekanaho, 2020; Hernandez, 2020). Benevolent sexism, which measures endorsement of traditional gender roles and belief in binary gender essentialism, is negatively correlated with acceptability ratings; but hostile sexism, which measures overt hostility towards women, has not been shown to be related (Bradley, 2020).

Beliefs about grammar and language change are among the most commonly cited metalinguistic reasons for disliking singular *they* (Bodine, 1975; Hekanaho, 2020). People who strongly endorse linguistic prescriptivism, which values "correct" language use and opposes changes to this standard, rate singular *they* as less acceptable (Bradley, 2020). Conversely, support of non-sexist language reform, which often focuses on avoiding masculine forms as the generic or the default, predicts higher acceptance of singular *they* (Hekanaho, 2020). Notably, predictors of attitudes about singular *they* now are similar to predictors of attitudes about non-

sexist language reform in the 1980s and 1990s. Alternatives to using *he* as the generic include phrases like *he/she* and *he or she*, as well as using singular *they* or *she* as a generic. People who were younger, who agreed that despite advances in equality women still face less-overt forms of sexism (Swim et al., 1995), and who had less conservative views about gender roles were more likely to support non-sexist language reform (Jacobson & Insko, 1985; McMinn et al., 1990; J. B. Parks & Roberton, 2004, 2008; Rubin et al., 1994; Sarrasin et al., 2012; Swim et al., 2004). I return to discussing the similarities between these two debates later in §0.8. While language attitudes certainly play a role, they do not explain all—or even, arguably, the majority—of variance in acceptability judgments about singular *they*. While the acceptability of generic, gender-unspecified forms of singular *they* was predicted by both linguistic prescriptivism and anti-trans prejudice, the acceptability of specific, gendered forms (e.g., coreferring with names) was primarily predicted by anti-trans prejudice (Hernandez, 2020).

Many of the predictors of attitudes about singular *they* are highly correlated. For example, people who are younger are more likely to be exposed to singular *they*, to know a TGD person, and to have less conservative beliefs about gender and sexuality (Minkin & Brown, 2021; Parker et al., 2019). It is still unclear whether age predicts attitudes towards singular *they* primarily by proxy through these other social factors, or if age-related differences in language learning and processing (DeDe & Knilans, 2016) are also at play. Some preliminary research has begun to try to tease these apart. Kramer et al. (2021), for example, argues that experience is a better predictor of using they/them pronouns correctly than age. Overall, the relationship between these predictors is still unclear and remains an area for future research.

0.4 Comprehension

Turning to comprehension, research has examined both offline comprehension—asking people about their interpretation of a sentence *after* reading or listening to it—and online comprehension—using measures such as reading time or eye gaze to observe people's

interpretation of a sentence *while* reading or listening to it (Allopenna et al., 1998; Liversedge & Findlay, 2000; Tanenhaus et al., 1995, 2000). Most studies of reading use one of two paradigms. In self-paced reading tasks, participants press a button to advance through the text one phrase or word at a time, measuring how long they spend at each region. In eyetracking while reading tasks, participants' eye movements are recorded while they read (relatively) naturalistically, measuring the sequence and duration of their saccades to (eye movements directed towards) and fixations on (gaze focused on) each word or phrase (Liversedge et al., 1998; Rayner & Juhasz, 2006). Generally, longer durations at and regressions back to a region are assumed to reflect having more difficulty understanding the text (Carpenter & Just, 1983; Henderson, 2013; Rayner, 1988). This is referred to as a processing cost or penalty. The following section is organized using the same three categories of singular *they* as §0.2.3: generic indefinite, specific definite gender-unspecified, and specific definite gender-specified (i.e., coreferring with proper names). But before discussing the existing research about comprehending singular *they*, I first review the prior research on gender agreement processing that it is based on.

0.4.1 Processing Gender Agreement

The majority of studies about gender agreement have used stereotype mismatch tasks, classically illustrated with:

> *[9] A man and his son were away for a trip. They were driving along the highway when they had a terrible accident. The man was killed outright but the son was alive, although badly injured. The son was rushed to the hospital and was to have an emergency operation. On entering the operating theatre, the surgeon looked at the boy, and said, "I can't do this operation. This boy is my son." How can this be?* (Sanford, 1985)

Many readers are confused by this story (see Morehouse et al., 2022 for a more recent overview). If, following the gender stereotype, the surgeon was assumed to be male, how could the boy's father be both dead in the accident and working at the hospital? It takes time to come to the

correct interpretation, that the surgeon is the boy's mother. In the basic mismatch paradigm, participants read a sentence where a character is first referred to by a stereotypically-gendered role noun (usually a profession) and later referred to with a pronoun that matches or mismatches the stereotype. Stereotype-mismatching pronouns (e.g., *surgeon…she*) take longer to read than stereotype-matching pronouns (e.g., *surgeon…he*), suggesting that when readers had first encountered the new referent, they used gender stereotype information to make an inference about the referent's gender (Kennison & Trofe, 2003; Reynolds et al., 2006). Upon reaching the mismatching pronoun, readers have to revise their interpretation, which incurs a processing cost. Alternatively, if a gender inference had not been made at the role noun, the process of making a gender inference at the pronoun would be the same regardless of stereotype match.

Converging evidence is also available from event-related potential (ERP) data, which uses electrodes placed on the scalp to measure changes in the brain's electrical field caused by neural activity (Samar, 2006). Stereotype-mismatching pronouns elicit a P600 effect: a positive wave in the centro-parietal regions, occurring 600–100ms after the word (Friederici, 2002). Some argue that the P600 indexes syntactic processing, and it can be observed when listeners encounter a syntactic error, a sentence with a complex syntactic structure, or new information that requires revising their interpretation of the sentence's syntactic structure (Hagoort et al., 1993; Kaan et al., 2000; Osterhout et al., 1994; Osterhout & Holcomb, 1992). Like a longer reading time, a P600 is interpreted as indicating that the sentence's syntactic structure is harder to understand. Notably, the P600 effect was still observed when participants judged the sentences with stereotype-mismatching pronouns to be acceptable (Osterhout et al., 1997; see Banaji & Hardin, 1996; Oakhill et al., 2005 for similar findings in a priming paradigm). Generally, the mismatch paradigm tends to remain agnostic about whether the mismatch effect arises more from gender stereotype beliefs or more from distributional knowledge. However, the fact that it occurs in the absence of negative metalinguistic judgments can be interpreted in favor of the gender mismatch being more due to statistical expectations in language processing (i.e., that *surgeon* occurs more frequently

with *he* than *she*) than to higher-level beliefs about gender (i.e., that surgeons should be men).

Grammatical gender languages provide stronger evidence for early, automatic inferences about the gender of referents when they are introduced into the discourse context. Carreiras et al. (1996) compared English results to Spanish, which marks gender on determiners and nouns. In English, the role noun provides stereotype information, and the referent's gender is not specified until the pronoun; in Spanish, the referent's gender is specified at the determiner:

[10] *1* *The footballer wanted to play in the match.*
2 *He$_{MASC}$/She$_{FEM}$ had been training very hard during the week.*

[11] *1* *El$_{MASC}$ futbolista/La$_{FEM}$ futbolista quería jugar el partido.*
2 *El$_{MASC}$/Ella$_{FEM}$ había estado entrenando mucho durante la semana.*

Because the determiner is gender-marked in Spanish [11], the referent's gender is disambiguated from the beginning. English [10] allows readers to make a guess about gender at the beginning using their knowledge about the gender-stereotyped role noun, but does not disambiguate the referent's gender until the pronoun in the second sentence, making it possible to have an incorrect interpretation of the referent's gender during the first sentence. Both English and Spanish showed a mismatch effect, where stereotype-mismatching sentences (***footballer…she*** in [10]) took longer to read than stereotype-matching sentences (***footballer…he*** in [10]), but it occurred at different points in the story. While English showed the mismatch effect at the pronoun (sentence 2 in [10]), Spanish showed the mismatch effect at the role noun (sentence 1 in [11]). This indicates that readers made inferences about the referent's gender as soon as gender information became available, regardless of whether or not it was necessary for comprehension.

An alternative explanation to these results, at least in English, is that the gender stereotype of the role noun is only activated when the pronoun is read, not when the role noun first introduces the referent. Thus, the slowdown would come only from incongruent gender cues, not from needing to revise an initial interpretation. In examples [10] and [11], this explanation would look like taking longer to process *she* coreferring with *footballer* because it does not match gender

stereotype predictions, but not because *the footballer* had been interpreted as male in the prior sentence. To rule out this explanation, Garnham et al. (2002) tested the mismatch effect with two probabilistic cues about gender: role nouns and either gendered clothing or physical traits:

[12] *1 The housekeeper/soldier*
2 was rushed to the hospital, taken to a private ward, and
3 gave birth half an hour later.

Here, both the role noun and the verb are stereotypically gendered, compared to prior experiments that paired a stereotypically-gendered role noun with a definitionally-gendered pronoun.[2] A mismatch effect will only occur if inferences about the referent's gender are made at both the role noun (line 1 in [12]) and at the stereotypically-gendered verb (line 3 in [12]), not just at the verb. In other words, *the soldier…gave birth* would not take longer to read than *the housekeeper…gave birth* unless the reader had already inferred that the soldier was male before reaching *gave birth*. Results showed mismatch effects for reading time and acceptability judgment reaction times, which indicates that definitionally-gendered language is not necessary to assign a gender to referents. Instead, comprehenders use probabilistic information such as gender stereotype and distributional knowledge about role nouns to make an inference about the referent's gender. This inference may be proven wrong after encountering further information, and needing to revise an interpretation of a referent who has been represented as (probably) one gender incurs a processing cost.

Another benefit of using grammatical gender languages to study agreement processing is that they allow grammatical and social gender cues to be manipulated separately, potentially providing conflicting information (Reali et al., 2015). Results from multiple languages suggest that

[2] *Arguably, pronouns are not definitionally gendered, since not all people who use she/her are women and not all women use she/her, and vice versa for men and he/him. However, treating pronouns as probabilistic cues—nearly all people who use she/her are women and nearly all women use she/her—would make the same predictions. Even if it is not a binary rule, the pronoun would still be a stronger cue about gender than the stereotypes associated with a role noun.*

stereotypical knowledge is activated independently from morphosyntactic cues (Molinaro et al., 2016; Schmitt et al., 2002). In addition to integrating potentially-conflicting gender cues, gender inferences need to be integrated with other cues about reference. Having the prior context indicate the gender of the referent eliminates the mismatch effect (Duffy & Keir, 2004; Kreiner et al., 2008):

[13] 1 *Jeff's/Lucy's power had been unreliable ever since the tornado. The electrician was a cautious man/woman who carefully secured his/her ladder to the side of the house before checking the roof. Jeff/Lucy suspected that high winds had loosened the connection to the power lines.*

2 *The electrician taught himself/herself a lot while fixing the problem.*

While the pronoun (*herself*) does mismatch the stereotypical role noun (*electrician*), it does not mismatch the interpretation supported by the preceding context (*the electrician was a cautious woman*). In stories like [13], *herself* does not take longer to read than *himself*, like it would if the second part were read in isolation. This does not imply that no difficulty integrating unexpected information happens, only that this process has already occurred before reaching the pronoun.

Gender information also interacts with broader discourse expectations. In English, pronouns are more likely to corefer with previously-mentioned names than upcoming new names. Participants took longer to reject gender-mismatching interpretations when a cue separate from gender—order of mention—supported the gender-mismatching interpretation (e.g., *John…she*) than when the order cue did not support the gender-mismatching interpretation either (e.g., *she…John*) (Kennison et al., 2009). Most of the studies discussed have presented stereotypical gender information when first introducing the referent, but stereotypical gender cues can also be used to revise prior interpretations made using other cues, such as order (Pyykkönen et al., 2010).

Even when it is a highly salient cue, gender information is not necessarily applied symmetrically. One experiment in German, aiming to distinguish between grammatical gender and stereotypical gender cues, measured reading times for pronouns in passages that described stereotypically-gendered jobs without using the grammatically-gendered role noun itself:

[14] *Original: L.K. vereinbart Termine, erledigt die Korrespondenz in einem Büro. Außerdem kann er eine fremde Sprache.*

Translated, maintaining the German word order: L.K. makes appointments, deals with the correspondence in an office. In addition speaks he a foreign language.

The mismatch effects for masculine pronouns coreferring with feminine jobs and for feminine pronouns coreferring with masculine jobs were different. In examples such as [14], masculine pronouns in a feminine context caused readers to look at the masculine pronoun for longer and to be more likely to look back at it, indicating that the gender mismatch effect occurred during both early and late stages of processing. In contrast, feminine pronouns in masculine contexts did elicit a mismatch effect, but only during late stages of processing. While female referents were interpreted to be compatible with both masculine and feminine contexts, male referents were only compatible with masculine contexts (Reali et al., 2015). This suggests that in the absence of definitive grammatical cues, gender stereotype information is interpreted more flexibly for women than men. Similarly, a priming task showed a slower reaction time for feminine-stereotyped role nouns followed by masculine pronouns, but no parallel effect for masculine-stereotyped role nouns followed by feminine pronouns (Cacciari & Padovani, 2007). This asymmetry likely reflects that changes in gender norms over the last 50 years have consisted more of women being able to do masculine things than men being able to do feminine things.

Finally, in some contexts comprehenders interpret gender-mismatched pronouns as referring to an unknown, upcoming referent, instead of trying to establish coreference with a referent that does not agree in gender. Hearing pronouns that mismatched definitionally-gendered antecedents (e.g., ***aunt…he***) elicited an Nref ERP response—a negative, left-frontal shift starting 300ms after the pronoun and continuing for the rest of the sentence (Nieuwland, 2014). The authors argue this indexes a high degree of referential ambiguity, where the listener fails to match the pronoun to a referent. In this example, the Nref occurs if the listener assumes that because *aunt* and *he* mismatch in gender, *he* refers to someone not mentioned yet.

0.4.2 Generic Indefinite Singular *They*

The studies discussed so far show that comprehenders automatically[3] and rapidly make inferences about gender when a new referent is mentioned—even when gender is irrelevant for comprehension—and that revising these gender inferences incurs processing costs. Turning to singular *they*, the earliest published data compared the processing cost incurred by generic indefinite singular *they* to those incurred by stereotype-mismatching *he* and *she* (Foertsch & Gernsbacher, 1997). Participants read "should" statements using an indefinite antecedent and a pronoun, then decided whether or not they agreed:

[15] *1 A nurse should have an understanding of how a medication works*
2 even if he/she/they will not have any say in prescribing it
3 because nurses must anticipate how a patient will respond to the medication.

The dependent measure was the time spent reading the clause including the pronoun (line 2 in [15]). There were four types of indefinite antecedents: stereotypically-masculine nouns, stereotypically-feminine nouns, gender-neutral nouns, and indefinite pronouns (e.g., *everyone*, *someone*); the four antecedents were crossed by three pronouns: *he*, *she*, and *they*. For gender-stereotyped antecedents, clauses with singular *they* were read more quickly than clauses with stereotype-mismatching pronouns and at the same speed as clauses with stereotype-matching pronouns. For gender-neutral antecedents, clauses with singular *they* were read at the same speed as clauses with *he* or *she*. For indefinite pronoun antecedents, clauses with singular *they* were read more quickly than clauses with *he* or *she*. The authors concluded that generic indefinite *they* was not more difficult to process than generic *he*, and in fact may be easier in some contexts.

---

[3] *I make a distinction between our higher-level beliefs about gender and our lower-level, automatic cognitive processing. The fact that comprehenders immediately infer a gender categorization for new referents reflects statistical knowledge about how language is used, but does not necessarily reflect a belief in gender as binary and transparent. For a theory of how automatic categorization processes interact with higher-level beliefs about gender in the context of making gender attributions, see Friedman (2014).*

An alternative baseline is a context where a plural mismatch effect is definitely expected. Instead of testing whether reading singular *they* is slower than *he* or *she* with singular antecedents like Foertsch & Gernsbacher (1997), Sanford & Filik (2007) tested if singular *they* retains the same plural mismatch effect as *he* and *she* with plural referents. Participants read sentences that included either a gender-unspecified indefinite antecedent (e.g., *a person, someone*) or a plural antecedent (e.g., *some people*), then either a singular pronoun (*him/her*) or a plural pronoun (*they*). This experiment employed an eyetracking while reading paradigm, which provides a more sensitive measure of processing disruptions than self-paced reading. The analysis included measures of early processing, including go-past reading time, which is the total time spent looking at the critical region when first encountering it; and measures of later processing, including total reading time, which is the total time spent looking at the critical region the first time and any times returning to it. Early processing measures showed a plural mismatch effect for *him/her* (e.g., *some people…him* > *a person…him*), but not for *their* (e.g., *a person…their* ≈ *some people…their*). The later processing measure showed plural mismatch effects for both *him*/*her* and *their*. However, the processing cost for singular *they* was small—only around 100ms. The authors interpret this to mean that singular pronouns show an earlier sensitivity to number mismatches than plural pronouns, but that when a mismatch is detected for singular *they*, it is easily accommodated.

0.4.3 Definite Singular *They*

The early studies about comprehension of indefinite singular *they* by Foertsch & Gernsbacher (1997) also tested specific definite forms, using similar “should” statements, e.g., [15] with *my nurse* instead of *a nurse.* For masculine- and feminine-stereotyped antecedents, clauses with singular *they* showed intermediate reading times: faster than clauses with stereotype-mismatching pronouns, but slower than clauses with stereotype-matching pronouns. For gender-neutral antecedents, there was no difference between singular *they* and *he*/*she*. Across antecedent definiteness and gender stereotypicality, singular *they* only showed a penalty

for definite, gender-stereotyped referents. Observing a processing cost for gender-stereotyped but not gender-neutral referents suggests that slower processing for definite specific singular *they* may be due more to expectations about when a speaker should know and mark the referent's gender, not necessarily due to number agreement (Foertsch & Gernsbacher, 1997).

A later study investigated reader's expectations about gender more directly (Doherty & Conklin, 2017). Participants read short stories that included a pronoun referring to a definitionally-gendered antecedent ("gender-known" condition, e.g., *spokeswoman*), a stereotypically-gendered antecedent ("high-expectancy" condition, e.g., *mechanic*), or a gender-neutral antecedent ("low-expectancy" condition, e.g., *cyclist*):

[16] *1 Adam recently had an accident in his car.*
2 He was emerging from a junction when he hit a cyclist and
3 knocked him/her/them
4 straight
5 off the bike.
6 Fortunately, the cyclist was not badly hurt.

Participants' eye movements were analyzed over three regions at and immediately after the pronoun. For gender-neutral antecedents, singular *they* was not read faster or slower than *he/she*. For stereotypically-gendered and definitionally-gendered antecedents, singular *they* did show costs, and critically, these patterns were different between the two antecedent types. At the verb + pronoun region (line 3 in [16]), singular *they* patterned with the gender-matching pronoun for definitionally-gendered antecedents (e.g., *spokeswoman*…*they* ≈ *spokeswoman*…*she*) but with the gender-mismatching pronoun for stereotypically-gendered antecedents (e.g., *mechanic*…*they* ≈ *mechanic*…*she*). At the adverb following the pronoun (line 4 in [16]), singular *they* was not faster or slower compared to gender-matching *he/she* for either antecedent type. At the final region (line 5 in [16]), singular *they* showed a cost compared to the gender-matching pronoun for definitionally-gendered, but not stereotypically-gendered antecedents. This is the reverse of the pattern seen at the verb + pronoun region, but now in the expected direction. However, Doherty

& Conklin note that the cost of singular *they* is much more transient and less robust than the cost of a gender mismatch (e.g., *spokeswoman…he*). The fact that stereotypically-gendered and definitionally-gendered antecedents behaved differently, both in the eyetracking data and in the acceptability judgment task that was used to pilot the stimuli, is evidence for probabilistic gender expectations in discourse processing. In a later study, late-intermediate and advanced learners of English showed similar patterns of reading times, adding further evidence indicating that late acquisition of singular *they* is feasible (Speyer & Schleef, 2019).

0.4.4 Singular *They* Coreferring with Proper Names

The findings so far show that when singular *they* refers to a specific definite—but gender-unspecified—referent, it does incur a processing cost, but one that is easily accommodated. The few studies about *they* coreferring with proper names (specific definite and gender-specified) also show a processing cost, but the magnitude and cause are still unclear. Early self-paced reading results have been inconclusive (Ackerman, 2018a, 2018b, 2020). An ERP study on people who used they/them pronouns for themselves or someone close to them—and thus were highly familiar with specific, gender-specified *they*—yielded mixed results (Prasad & Morris, 2020). Participants showed no P600 effects, which index agreement and syntactic violations, for *themselves* coreferring with specific indefinite (e.g., *someone*) and specific definite gender-unspecified (e.g., *the participant*) referents, but did show P600 effects for specific definite gender-specified referents (e.g., proper names).

More recent preliminary findings showed that *they* coreferring with proper names elicited a P600 effect like gender-mismatching *he* and *she* (e.g., *Mary…he*), but only mismatching *he* and *she* elicited an Nref effect. This suggests that singular *they* is causing some degree of grammatical processing difficulty for listeners, but that it is not causing referential failure like strong gender mismatches (Chen et al., 2023). A related study used a maze task, which presents participants with a sentence one word at a time, and at each word, they have to pick which of two words could

continue the sentence grammatically. At the critical location, participants have to decide between a pronoun (*he*, *she*, plural *they*, or singular *they*) and a different part of speech that is not grammatical at that location, e.g., *Mary watched TV before bed because* [*they* or *liked*], where a pronoun but not a verb can follow *because*. Participants were slower to decide that singular *they* could continue the sentence compared to gender-matching *he/she*, but not as slow as with gender-mismatching *he/she* (e.g., *Mary*…*she* < *Mary*…*they* < *Mary*…*he*) (Shenkar et al., 2023).

Finally, two sets of studies have investigated how comprehenders use discourse context to understand singular *they*, one using an offline comprehension question and one using mouse tracking. In the offline comprehension task (Arnold et al., 2021), participants learned about three characters (1 he/him, 1 she/her, and 1 they/them), then read two-sentence stories with comprehension questions probing which character(s) they interpreted the pronoun as referring to. In the training phase, participants read stories with only one character, which strongly supported interpreting *they* as singular, e.g., [17]. In addition to seeing the character's pronouns modeled in the stories, some participants were given more direct information, e.g., *Alex uses they/them pronouns*. The majority of participants endorsed the singular interpretation (Alex is the one who fell down) during the training phase, and only participants who interpreted *they* as singular in all training trials were included in the primary analysis. After confirming that participants understood that the character uses they/them pronouns, the test phase investigated how *they* was interpreted in stories with two characters, where *they* could plausibly be singular or plural, e.g., [18].

[17] 1-character context:
"Alex went running. They fell down." Who fell down?

[18] 2-character context, named first:
"Alex went running with Liz. They fell down." Who fell down?

[19] 2-character context, named second:
"Liz went running with Alex. They fell down." Who fell down?

Two factors increased the proportion of singular interpretations: direct information and discourse

context. Among participants who had always chosen the singular interpretation in the one-character training trials, participants who had also been directly told that the character uses they/them pronouns were more likely to choose the singular interpretation in the two-character context. This demonstrates that people can learn about someone's pronouns and use that information to disambiguate singular from plural *they*. Participants were also more likely to choose the singular interpretation when the they/them character was named first [18] than when the they/them character was named second [19], which replicates prior findings for when *he* or *she* is ambiguous between two referents (Arnold et al., 2000, 2007; Brown-Schmidt & Toscano, 2017; Gernsbacher, 1989; see §4.1).

In the online comprehension task (Arnold et al., 2023), participants listened to stories about pairs of the characters interacting with objects:

> *[20] Liz and Alex were cleaning up after a dinner party. Liz handed a towel to Alex. Then they dried the plates.*

In the screen corresponding to [20], Liz is pictured with pots and pans, Alex is pictured with plates and silverware, and Alex and Liz together are pictured with cups and bowls. The participants' task was to click on the objects that the character in the final sentence was using, and their reaction time and the number of times their mouse movement changed direction horizontally (i.e., moving back and forth between options) were recorded. If participants understand that *they* is singular and refers to Alex, they would be able to anticipate which object to click on and respond faster after hearing *plates*. Unsurprisingly, participants were slower to click on the image for stories using singular *they* than *he* and *she*. However, the degree of processing cost was modulated by discourse context, like in the offline comprehension task. When the they/them character was mentioned first and was thus more likely to be the antecedent of the pronoun, the number of mouse direction shifts was not increased compared to *he* and *she*. The authors interpret this to mean that when the discourse context supports interpreting *they* as singular, listeners still

understand it a bit more slowly, but are not experiencing additional competition between the singular and plural interpretations (i.e., moving their mouse between the images for Alex and Alex & Liz). These two experiments provide further evidence that any processing costs for gender-specified singular *they* are manageable for comprehenders. Moreover, demonstrating that the character being named first instead of second supports comprehending *they* as singular suggests that people are using similar cues to disambiguate between singular and plural *they* as when other pronouns are ambiguous. If this is the case, it points to successful integration of singular *they* into existing reference resolution mechanisms, instead of treating it as an exception.

0.5 Production

One line of research about the production of singular *they* situates it in sociolinguistic theories of politeness. Conrod (2019b) argues that gendered pronouns function much like honorific markers, reflecting and negotiating a social relationship, as opposed to encoding a static grammatical feature or making a claim about some biological reality. Singular *they* introduces additional options for speakers—including the gender-unspecified form for people who use he/him or she/her—and so the choice between these options carries information. Speakers can take a variety of approaches to this decision. Someone may prioritize quantity (including enough information) over relevance and quality (including only relevant and accurate information) (Grice, 1975). In this case, the speaker could choose to be as specific as possible and use *he* or *she*, even when gender information is not relevant, or their use of *he* or *she* may be incorrect (e.g., the exchange in [5]). If, on the other hand, the speaker prioritizes relevance over quantity, they could choose to be more vague and use *they*, even if they could include information about the person's gender. Similarly, if a speaker prioritizes quality over quantity, they could choose to use *they* when the person's gender is unknown, instead of making an inference about whether *he* or *she* would be correct (e.g., the exchange in [4]).

In addition to communicating information, conversations are also building social relationships. The politeness theory framework proposes that people need to maintain both positive face (being liked and confident) and negative face (being free from obligations and impositions), and what they say can either maintain or threaten their face (Brown & Levinson, 1987). Conrod (2019b) describes several face constraints for gendered pronouns: (1) *don't ungender*, because not specifying gender is an imposition on someone's positive face; (2) *don't misgender,* because specifying the wrong gender is an imposition on someone's positive face; and (3) *don't correct me*, because telling someone which pronouns to use is an imposition on their negative face. When a speaker does not know which pronouns to use, they can choose to prioritize (1) by making a guess, or to prioritize (2) by using gender-unspecified *they* or by avoiding pronouns. Much of the debate around pronouns is a conflict between (2) and (3), where people argue that correcting someone threatens face more than misgendering someone. Examples of this reasoning are discussed later in §0.8. Finally, choosing pronouns involves taking an affective and evaluative stance about both the person you are speaking about and the person you are speaking to (Du Bois, 2007). This stance-taking may be more overt, such as when speakers intentionally misgender someone to express disapproval and disagreement, or more subtle, such as when speakers leave gender unspecified to imply social distance (e.g., Conrod, 2019a).

Only a few experiments have tested language production. Generally, their preliminary findings indicate that speakers can produce specific definite singular *they* in both writing (Kramer et al., 2021) and speech (Arnold, Venkatesh, et al., 2022), especially when shown examples of the person being referred to with they/them. Production appears to cluster into the same three patterns as acceptability judgments (Camilliere et al., 2021; Konnelly & Cowper, 2020), with participants either not using singular *they* at all, using it only for referents introduced by last name (such that gender was unspecified), or using it for referents introduced with last names or first names (such that gender was inferrable) (Kaiser & Post, 2023).

In addition to how speakers can learn to produce they/them pronouns for a person instead of he/him or she/her, another open question is how speakers integrate singular *they* into their production system when deciding to use a pronoun instead of a different type of referring expression. Since pronouns are the most reduced forms compared to proper names and noun phrases, speakers only tend to use pronouns when the referent is highly salient, in focus, or accessible. In other words, speakers tend to choose the option that includes the least information if they think the comprehender already has enough information to identify the referent (Arnold & Zerkle, 2019). The counter-argument to singular *they* being too ambiguous to understand is that standard pronouns like *he*, *she*, and plural *they* frequently have the potential to be ambiguous. Speakers already have strategies to avoid too much ambiguity, and comprehenders already have strategies to interpret ambiguous pronouns. If specific singular *they* is being successfully integrated into our language use, then speakers will use or adapt existing strategies to minimize ambiguity. In a corpus study of news articles about nonbinary people who use they/them pronouns, writers were less likely to use pronouns overall for people who use they/them compared to people who use he/him or she/her. However, this difference was not explained by potential ambiguity, which the authors interpret to mean that writers may have been avoiding pronouns that were unfamiliar for them or their intended readers, not that a referent using they/them created contexts where a pronoun would have been uniquely ambiguous (Arnold, Marquez, et al., 2022). In a speech production study, participants were less likely to use a pronoun to continue a story that had introduced two characters than a story that had introduced only one character. Instead, people were more likely to use a name in this context since a pronoun could refer to two possible referents. However, pronoun use was reduced to the same degree for *they* compared to *he* and *she*, again suggesting that singular *they* does not introduce a unique level of ambiguity (Arnold, Venkatesh, et al., 2022).

0.6 Frequency

As society has become more accepting of LGBTQ+ identities, the TGD community in the United States has grown in both size and visibility. A meta-analysis of population-based, probability samples from 2006–2016 estimated that 1 in 250 U.S. adults are transgender (Meerwijk & Sevelius, 2017). In 2014, the CDC Behavioral Risk Factor Surveillance System Survey—a representative, probability-based survey about the health behaviors and conditions of U.S. adults—added a set of demographic questions about sexuality and gender, which around 30 states now include. An average of 1 in 200 people have said they considered themself transgender (CDC, 2014, 2015; analyzed in Crissman et al., 2017; Herman et al., 2022), with the rate rising to 1 in 128 in the most recent survey (CDC, 2021). The TGD community skews younger, even more so than the broader LGBTQ+ community (Harrison et al., 2012; Herman et al., 2022; James et al., 2016; Streed Jr. et al., 2017). While the sample sizes of population-based surveys of teens are smaller than those of adults, a larger proportion of U.S. teens identify as[4] transgender—between 1 in 50 and 1 in 75 (CDC, 2019; Eisenberg et al., 2017; Herman et al., 2022; Kidd et al., 2021; Perez-Brumer et al., 2017; Shields et al., 2013; Twenge, 2023).

An increasing proportion of the transgender community is nonbinary: In 2014, 17% of transgender respondents in the CDC BRFSS survey did not identify as men or women; in 2021, this increased to 33% (CDC, 2014, 2021; similar or higher rates in Grant et al., 2011; Harrison et al., 2012; James et al., 2016; C. Parks et al., 2023; UK Government Equalities Office, 2018; Wilson & Meyer, 2021). The majority of the population-based surveys discussed so far ask if the person identifies as transgender, and then if so, if they are a man, a woman, or gender-

[4] *To be clear, I use language like "identify as" and "consider themself" not to imply that people's identities are less real, but to describe surveys asking people about the labels they use. Decisions about how to describe gender and sexuality are complicated and personal, and I want to maintain a distinction between the set of people who* do *say "yes, I am X," and the set of people who fall under the definition of X and* could *say "yes, I am X" or may be grouped as X in research analyses. Often in LGBTQ+ research the set of people who could identify as X is larger than the set of people who do identify as X, which makes estimating and talking about population sizes tricky.*

nonconforming/something else. Since not all nonbinary people also identify as trans, the size of the broader TGD community is likely still underestimated (James et al., 2016; Meerwijk & Sevelius, 2017; Wilson & Meyer, 2021). The U.S. Household Pulse survey, which asks American adults every few weeks about employment, living, and health conditions, added a survey question about gender in 2021. The response options include *male, female, transgender*, and *none of these*. While this is not the ideal way of asking about gender (NASEM, 2022), the dataset is large enough that it is able to start correcting for the undercount of nonbinary people. Rates of nonbinary identity are increasing particularly for Gen Z adults (ages 18–26 in 2021 and 2022): 2.3% identify as trans and 3.3% as "none of these." To put this into perspective, out of the 39 million Gen Z adults in the U.S., about 2 million are trans and/or nonbinary, which is more than the population of Phoenix, AZ—the fifth-largest city (Twenge, 2023). Analyzing changes in the Household Pulse survey over the past few years, Twenge argues that Gen Z represents a massive, ongoing generational shift in views about gender (2023).

While the changes are largest in younger generations, support for and familiarity with TGD people has been growing across generational, political, and religious groups. Currently, 40% of U.S. adults know a trans person and 16% know a nonbinary person, and the proportion who have a trans family member or close friend has more than doubled from 11% in 2011 to 24% in 2019 (Jones et al., 2019; Minkin & Brown, 2021; PRRI, 2021). The majority (70%) say they would be somewhat or very comfortable learning that a friend, coworker, or community member was trans, and a slim majority (55%) say the same for their child (PRRI, 2021). On average, 60% of adults said they became more supportive of transgender rights between 2012 and 2019, compared to 25% who said they became less supportive. 40% said that trans people face "a lot of stigma or negative social judgment in their community" (Jones et al., 2019). Among young adults specifically (ages 15–24), the majority (80%) said that there was a lot of discrimination against trans people, and only 12% perceived it as decreasing during 2017 (Jones et al., 2018).

Within the growing TGD community, around 80% of nonbinary people and 50% of trans people say that they use they/them pronouns (Cassian, 2021, 2022; Cheung et al., 2020; C. Parks et al., 2023). This corresponds to rapid increases in familiarity with singular *they* in recent years: Around 70% of people under 40, 60% of people 40–55, and just under 50% of people 55+ say they have heard a little or a lot about "people preferring that others use gender-neutral pronouns such as *they* instead of *he* or *she* when referring to them" (Parker et al., 2019). 46% of people ages 18–29, 29% of people 30–49, 18% of people 50–64, and 11% of people 65+ personally know someone who prefers gender-neutral pronouns in 2021, up from 8–32% in 2018.[5] Correspondingly, about 60% of people 18–29, 50% of people 30–64, and 40% of people 65+ say they would be somewhat or very comfortable using gender-neutral pronouns—which would most likely be singular *they*—for someone who asked. Political affiliation is a stronger predictor than age here, with 40% of Democrats saying they would be very comfortable and 40% of Republicans saying they would be very uncomfortable (Minkin & Brown, 2021). Compared to gender-neutral pronoun use, U.S. adults are more divided about the concept of the gender binary, with 40% saying they somewhat or strongly believe there are a range of genders and 60% saying they somewhat or strongly believe there are only two (Jones et al., 2019; PRRI, 2021; Twenge, 2023).

0.7 Misgendering

Despite these increases in visibility, TGD people in the U.S. continue to face high rates of discrimination and harassment in workplaces, schools, healthcare centers, and public spaces (Cruz, 2014; Grant et al., 2011; James et al., 2016; UK Government Equalities Office, 2018); and are more likely to struggle with depression, anxiety, suicidality, and overall poor health than their

[5] *Knowing a person who uses gender-neutral pronouns is more common than knowing a nonbinary person (Minkin & Brown, 2021; PRRI, 2021), but this is not particularly surprising given that not all people who use gender-neutral pronouns are nonbinary, and that people may tell others what gendered language they prefer, without getting into the details of their gender identity.*

cisgender LGBQ+ peers and the general population (Bockting et al., 2013; Chodzen et al., 2019; Cruz, 2014; Eisenberg et al., 2017; Kattari et al., 2019; Perez-Brumer et al., 2017; Streed Jr. et al., 2017). For nonbinary people specifically, health disparities and experiences of discrimination tend to be similar, or even worse, than other groups in the TGD community (Budge et al., 2014; Burgwal et al., 2019; Butler et al., 2019; Cheung et al., 2020; Clark et al., 2018; Harrison et al., 2012; James et al., 2016; Kattari et al., 2020; Keuroghlian et al., 2015; Lagos, 2018; Lefevor et al., 2019; Morris & Galupo, 2019; Newcomb et al., 2020; Sterzing et al., 2017; E. A. Tebbe & Moradi, 2016; Veale, Watson, et al., 2017). Recent work has contextualized these experiences in the minority stress framework, where health disparities in a marginalized group arise from the array of stressors caused by societal stigma against them (Chodzen et al., 2019; Howe, 2019; Lindley & Galupo, 2020; Puckett et al., 2020; Rood et al., 2016; E. A. Tebbe & Moradi, 2016; Testa et al., 2015; Valentine & Shipherd, 2018; White Hughto et al., 2015).

Misgendering—where someone is referred to using gendered language that does not match their identity—is a commonly reported minority stressor affecting the TGD community. TGD people describe being misgendered as alienating, devaluing, invalidating, and painful (Cordoba, 2020; Goldberg et al., 2019; Gunn, 2020; A. Johnson, 2019; Pitcher, 2017; Saltzburg & Davis, 2010; Truszczynski et al., 2020). This is because the language we use to refer to ourselves—names, pronouns, kinship roles, community affiliations—creates and reflects parts of our identity:

> *If, as posited by social constructionism, language constitutes and creates the meaning of our lives, then not having language to fit how [genderqueer youth] view themselves leaves them feeling outside of and "not counted" in the human experience...It seems that in trying to fit gender variant young people into a languaged existence that does not correspond to how they view themselves, society deprives them of sense of self, self-value, and a recognized social existence.* — Saltzburg & Davis (2010)

As a consequence of having their identities forgotten, contested, or overwritten, people who are misgendered more frequently show higher rates of depression, stress, suicidality, disordered

eating, body dissatisfaction, and low self-esteem (Galupo et al., 2020; McLemore, 2015, 2018; Mitchell et al., 2021; The Trevor Project, 2020). Using gender-affirming pronouns, names, and body terms is particularly important in healthcare contexts, as it forms a key part of how positive or negative the experience is for TGD patients (Baldwin et al., 2018).

Nonbinary TGD people are misgendered frequently: only 1 in 10 nonbinary youth report all or most people respecting their pronouns (The Trevor Project, 2020), and 2 in 3 nonbinary people did not ask for correct pronouns at work to avoid discrimination (James et al., 2016). Additionally, some studies report that getting people to use new pronouns is more difficult than getting people to use a new name (Barbee & Schrock, 2019). While most studies about misgendering report participants' genders, but not the gendered language they use, the fact that a large proportion of nonbinary people use they/them pronouns (Cassian, 2021, 2022; Cheung et al., 2020; C. Parks et al., 2023)—combined with the relative newness of specific singular *they*—means we can infer that people who use they/them pronouns are likely misgendered regularly.

Using they/them pronouns often places TGD people in a double bind, where they are forced to choose between experiencing more discrimination by openly identifying as TGD, or between getting misgendered. For example, when searching for jobs, the majority of nonbinary people said that identifying themselves as nonbinary to a potential employer would hurt their job search somewhat (60%) or very much (25%). These expectations were confirmed by an experiment that measured how adding they/them pronouns to a resume affected application outcomes. The otherwise-identical resume received 9% more interest, and when asked directly about their impressions, hiring managers rated the applicant with a resume including they/them pronouns as 7% less qualified and were 4% less likely to invite them for an additional interview (McGonagill, 2023).

Misgendering is a form of microaggression (Chang & Chung, 2015; Nadal, 2019; Nadal et al., 2014, 2016). In the original microaggression framework, developed to describe experiences with racism, microaggressions fall into three categories: microinsults, which carry subtle or implied

snubs (i.e., that the speaker did not expect the person to be as competent as they are); microassaults, which insult someone through name-calling or avoidant behavior; and microinvalidations, which deny the experience, beliefs, or emotions of the person (i.e., "but I don't see color" as negating someone's experience of how their race affects how they are treated) (Sue et al., 2007). Intentional misgendering—where someone deliberately uses the wrong name or pronouns to insult someone or deny the legitimacy of their identity—would fall under the category of microassaults (Conrod, 2019a; Nadal, 2013). Unintentional misgendering—where someone uses the wrong name or pronouns because of an incorrect inference about someone they don't know (i.e., that all people they perceive as feminine use she/her) or because of a speech error after someone changes names or pronouns—would fall more under the category of microinvalidations (Gunn, 2020; McLemore, 2015). The key idea of the microaggressions framework is that while many individual events may be minor or perpetrated unintentionally, the toll of experiencing them frequently is substantial and cumulative. Nonbinary people described their reactions to being misgendered using similar distinctions as the microaggressions framework, where both the perceived intent and the proximity of the speaker mattered (Cordoba, 2020). Being misgendered by people they were close to was interpreted as more intentional and experienced as more hurtful. While being misgendered by acquaintances or strangers was interpreted as less intentional and hurt less, these negative experiences accumulate constantly. One genderqueer person described their experience as:

> *But sometimes it just feels like a thousand paper cuts. And you know, by the end of the day you're like, "how many times has someone misgendered me today?" It's really difficult to speak about someone without using a pronoun, particularly when you're not aware that's a problem. And so, sometimes by the end of the day you just feel like, you've just been rubbed raw and you're like, "How many billions of times did someone refer to me as* she *and* Ms. *today?"*
> — interviewed in Cordoba (2020)

Other work has described how social support can mitigate the negative mental health effects of

misgendering and other transphobic microaggressions (Bockting et al., 2013; McLemore, 2015; Penklis, 2020; Puckett et al., 2019; Testa et al., 2015; The Trevor Project, 2020; Veale, Peter, et al., 2017; Weinhardt et al., 2019). Little research, however, has focused on preventative strategies. Reducing how often TGD people are misgendered is one avenue for ameliorating disparities in mental health.

0.8 Sociopolitical Context

Like other language innovations in the LGBTQ+ community, singular *they* evokes overt hostility about nonbinary and transgender identities, language change, and people's right to determine how they are referred to. More broadly, misgendering and language reform have become a prominent topic in the "political correctness" and "wokeness" wars. A *Wall Street Journal* op-ed by journalist and political speechwriter Peggy Noonan expresses the major themes of opposition:

> *I've been thinking about the language and behavioral directives that have been coming at us from the social and sexual justice warriors who are renaming things and attempting to control the language in America. In one way it's the nonsense we've all grown used to, but it should be said that there's an aspect of self-infatuation, of arrogance, in telling people they must reorder the common language to suit your ideological preferences. There is something mad in thinking you should control the names of things...Offices and schools are forced to grapple with all the new gender-neutral pronouns…Use "they" a lot. It's gender neutral…This is grammatically incorrect but so what? Correct grammar, and the intelligibility it allows, is a small price to pay for inclusion and equality. We are being asked to memorize all this, to change hundreds of years of grammar and usage, to accommodate the needs or demands of a group that perceives itself as beleaguered.* — "What were Robespierre's Pronouns?" (2019)

On its face, the opposition to singular *they* is about grammatical correctness and clarity, presuming that English grammar exists in some pure form which must be guarded and preserved. While some people claim that "correct" grammar cannot change, others argue that change is simply too difficult a request. For people like Noonan, the costs of changing language use are

massive, and the costs of misgendering others are grossly exaggerated.[6] Arguments about grammar frequently mask arguments about gender: are nonbinary identities real and important enough to warrant consideration in our language use? Noonan implies that they are not when she dismisses TGD people as "a group that perceives itself as beleaguered." Former academic social psychologist turned conservative influencer Jordan Peterson is more overt about opposing TGD identities more than he supports traditional grammar:

> *I don't recognize another person's right to decide what words I'm going to use, especially when the words they want me to use, first of all, are non-standard elements of the English language and they are constructs of a small coterie of ideologically motivated people. They might have a point but I'm not going to say their words for them…There's not enough evidence to make the case that gender identity and biological sexuality are independently varying constructs. I don't believe that it's reasonable for our society to undermine the entire concept of binary gender in order to hypothetically accommodate a tiny minority of people.*
> — "I'm not a bigot" Meet the U of T prof who refuses to use genderless pronouns (2016)

I have already presented population data showing that the TGD community is not so tiny of a minority that they are not, practically speaking, worth consideration (§0.6). More importantly, the clinical evidence shows that misgendering causes harm (§0.7). One of the premises of this dissertation is that getting someone's pronouns right is not a unique or unreasonable demand. Instead, I place pronouns in the same category as names and titles, where we generally *do* agree that people get to "decide what words [other people] are going to use." In some cases, it may be linguistically difficult, like how someone's name may be hard to pronounce for non-native speakers, but doing our best to get it right is part of being polite.

While the debate about singular *they* and nonbinary identities is relatively recent, the broader debate about gendered language, sociopolitical influences on language, and the standards of correct speech has been happening for centuries (Bodine, 1975). The following 1880

---

[6] *See Conrod (2019b)'s analysis of politeness and face constraints, discussed in §0.5.*

complaint about generic *they* as an alternative to generic *he* bears striking similarity to Noonan's complaints about singular *they*:

> Their *is very commonly misused with reference to a singular noun. Even John Ruskin has written such a sentence as this: "But if a customer wishes you to injure their foot or to disfigure it, you are to refuse their pleasure." How Mr. Ruskin could have written such a sentence as that (for plainly there is no slip of the pen or result of imperfect interlinear correction in it), or how, it having been written, it could be passed by an intelligent proof-reader, I cannot surmise. It is, perhaps, an exemplification of the straits to which we are driven by the lack of a pronoun of common gender meaning both* he *and* she*,* his *and* her*. But, admitting this lack, the fact remains that* his *is the representative pronoun, as mankind includes both men and women.* — *Everyday English* (1880)

Complaints about singular *they* and *he or she* as alternatives to generic *he* were essentially the same: they were argued to be grammatically incorrect and unclear, despite common usage (Bodine, 1975). More critically, non-sexist language reform was considered unnecessary (Blaubergs, 1980; Martyna, 1980; J. B. Parks & Roberton, 1998). Women's opposition to using generic masculine language was trivial and unreasonable, a statement that is far more political than linguistic (Ehrlich & King, 1992; Penelope, 1982). In each of these arguments, the underlying assumption is that the status quo and defense of it is not ideologically motivated, but support for language change is.

This dissertation takes it for granted that people asking to have their identities respected and to not be misgendered is not, as the subtitle of Noonan's essay claims, "the work of sociopaths who politicize language" (2019). Unfortunately, pronouns have become an overt symbol of whether or not you acknowledge that trans and gender diverse people exist, with "using pronouns" escalating from a dog whistle to one of the focuses of current anti-LGBTQ+ legislation. 2023 has already seen a record number of bills introduced in U.S. state legislatures targeting transgender rights (Human Rights Campaign, 2023). Many of these bills include policies about misgendering—alongside more extreme policies such as outlawing all gender-affirming medical care (ACLU, 2023). The majority of the current bills focus on public school contexts and minors.

Multiple states have successfully banned (TN S.B. 0466/H.B. 1269, 2023) or attempted to ban (AZ S.B. 1001, 2022; IN H.B. 1346, 2023; ND S.B. 2231, 2023)[7] public schools from requiring that employees refer to a student using pronouns different than he/him or she/her corresponding to the student's sex assigned at birth. The proposed legislation frequently moves beyond banning consequences for misgendering to banning the promotion of *not* misgendering (IN H.B. 1346, 2023). For example, one Oklahoma bill bans "creating, enforcing, or endorsing a policy that respects, favors, endorses, or promotes non-secular self-asserted sex-based identity narratives or sexual orientation orthodoxy." This includes "mandating non-obvious pronoun changes that respect gender identity ideology and sexual orientation orthodoxy" because these are "naked assertions that are implicitly religious and have a tendency to erode community standards of decency and promote licentiousness" (OK S.B. 937, 2023).[8]

The prohibition of promoting not misgendering often moves to creating consequences for not misgendering. In most variations of this bill, school employees would not be allowed to use a student's chosen pronouns unless the school has written permission from a parent (IN H.B. 1346, 2023; NH H.B. 619, 2023; OK S.B. 30, 2022; AZ S.B. 1001, 2022; ND S.B. 2231, 2023; ND S.B. 2260, 2023),[9] as well as medical documentation for diagnosed gender dysphoria (IN H.B. 1346, 2023). This would make it difficult for TGD students to be out at school without proactive support from a parent, which many teenagers do not have (Baum et al., 2012; The Trevor Project, 2020, 2022). Instead, school employees would be legally required to notify parents if student uses "expresses or indicates a desire to change…(A) name; (B) attire; or (C) pronoun, title or word to identify the student; in a manner that is inconsistent with the student's biological sex assigned at

---

[7] *As of May 30, 2023, the Indiana bill is dead, the Arizona and North Dakota bills were vetoed by the governor, and the Tennessee bill was signed into law.*

[8] *Advancing as of May 30, 2023.*

[9] *As of May 23, 2023, the Indiana and New Hampshire bills are dead, the Oklahoma bill is advancing, and the Arizona and North Dakota bills were vetoed by the governor.*

birth" (IN S.B. 354, 2023).[10] In 2022, only half of TGD youth said their school was supportive, but only a third said their home was supportive (The Trevor Project, 2022). This indicates that while schools are not supportive enough overall, they are still an environment where many TGD youth receive support that they are not otherwise receiving at home—a possibility that the current legislation aims to remove.

Anti-LGBTQ legislation also targets education more broadly, not just policies about specific transgender students. Schools would not be allowed to provide instruction "recognizing expressed gender" in classrooms or in professional development (ND S.B. 2231, 2023), or to teach "different pronouns other than those in common use in the English language when referring to the male or female" (NH H.B. 619, 2023). Books that "promote gender fluidity or gender pronouns" would be in the category of books that parents could report and have removed from schools (AZ S.B. 1700, 2023).[11]

Why talk about pronouns now? Given how legal protections and healthcare access are being eroded in the U.S., misgendering is far from the most immediate or severe hardship facing trans and gender diverse people. A similar argument was made in the non-sexist language debates (Blaubergs, 1980; J. B. Parks & Roberton, 1998), and probably comes up in every language reform effort. Claiming that misgendering is worth addressing is not claiming that language is the worst problem, and a form of injustice does not need to be one of the most severe to be worth fighting. Much as we might like to ignore gendered language in favor of other, bigger problems—or at least people who aren't linguists might like to think less about language—English does not give us the option of ignoring it. Because Standard English requires us to encode gender in many of the pronouns, titles, and other terms we use to refer to people, gender is part of

---

[10] *Dead as of May 23, 2023.*

[11] *As of May 23, 2023, the North Dakota bills are vetoed and not advanced to a full vote, respectively. The Arizona bill has passed the senate and is being considered by the house.*

“thinking for speaking” (Slobin, 1996), where we have to think about and assert gender even in contexts where we might otherwise prefer not to. In the words of one genderqueer trans person:

> *I think gender becomes very prominent when you are constantly being misgendered, and living with dysphoria. I don't necessarily want it to be important, but it is in the way I guess an injury is important, in that you have to be careful with it, pay it more mind than a non-injured part. [...] I really wish it didn't have to be, wish it wasn't enshrined in our language that people's gender is the most important thing and the first thing we need to know, even for a casual “hello.”*
> — interviewed in Cordoba (2020)

A second reason that misgendering is important to address now is that it is a tractable and winnable problem. The debates about misgendering are headed by a small but vocal minority. As I discussed in §0.6, the majority of people are supportive and open to learning. Critically, misgendering is one aspect that individual allies have the ability to concretely affect. While most people cannot, for example, change healthcare policy in their state, they can change their own language and the norms of their communities, making their environments more welcoming to the trans and gender diverse people in their communities. Singular *they* is a case where linguistics and cognitive psychology have relevant expertise and an opportunity to contribute to socially relevant research.

CHAPTER 1

Experiment 1: Associating Pronouns with a Person

1.1 Motivation

In addition to learning that the grammatical representation of *they* allows it to corefer with specific singular referents, learning to use they/them pronouns may require a change in how speakers access pronouns during language production. Extending existing models of pronoun production from grammatical gender to social gender (Ackerman, 2019; McConnell-Ginet, 2014) would predict that pronouns are accessed based on morphosyntactic gender marking associated with a person's name (e.g., Schmitt et al., 1999) or based on semantic/conceptual features of a person (e.g., Antón-Méndez, 2010). In order to use they/them for a person instead of the expected he/him or she/her, speakers may instead need to recall episodic information about the person's stated pronouns or which pronouns other speakers use to refer to them. It is clear from other contexts, such as referring to pets with *he* and *she*, that people can learn which gender-marked pronouns to use in contexts with few gender cues based on name or appearance. This suggests that learning specific singular *they* should be feasible, but may involve retrieving a person's pronouns from memory, rather than inferring them based on cues such as their name.

The first experiment investigated how people learn to associate pronouns with a person when someone's pronouns cannot be inferred from the gender association of their first name. Participants read a series of vignettes which introduced 12 characters, each of whom was associated with a name, pronouns (he/him, she/her, or they/them), a pet, and a job. Memory for pronouns was tested in a multiple-choice recognition task, and production of pronouns was tested in a written sentence completion task.

1.2 Experiment 1A

The design and analysis plan were preregistered on the Open Science Framework (10.17605/OSF.IO/CMKW5). Materials, de-identified data, and analysis code are available at this dissertation's Github repository.

1.2.1 Methods

Participants. 102 undergraduate students from Vanderbilt University completed the study for partial course credit. The study was conducted online and took approximately 15 minutes. To characterize the participant sample, they were asked about their age (M_{age} = 19.49, SD_{age} = 1.38); gender (79 female, 22 male, 1 nonbinary), and English experience (86 "native (learned from birth)," 15 "fully competent in speaking, listening, reading, and writing, but not native," and 1 "limited but adequate competence in speaking, reading, and writing"). Following best practices for TGD-inclusive study design, the question about gender was a free-response box (Ansara & Hegarty, 2014; Cameron & Stinson, 2019; NASEM, 2022; Zimman, 2017) (§B.1.1).

Materials & Procedure. Participants were introduced to 12 characters, each of whom had 4 associated facts: name, pronouns, job, and pet. There were 6 typically-masculine names (Andrew, Brian, Daniel, James, Kevin, Michael) and 6 typically-feminine names (Amanda, Emily, Jessica, Laura, Melissa, Stephanie), such that 4 characters had he/him pronouns and masculine names, 4 characters had she/her pronouns and feminine names, 2 characters had they/them pronouns and masculine names, and 2 characters had they/them pronouns and feminine names. While people who use they/them pronouns have a variety of names, the choice of gendered names here means that whether a character used the gendered pronouns associated with their name or they/them could not be predicted on the basis of the name. Thus, participants had to learn each name-pronoun pairing. Characters also had 1 of 12 jobs and 1 of 3 pets. The pet fact had the same distributional characteristics as the pronouns, but was not associated with prior expectations based on the name or with strong sociopolitical beliefs, making pet learning a

reference point for comparison. Participants were randomly assigned to 1 of 3 lists. Name-pronoun pairs were counterbalanced such that each name appeared with the expected he/him or she/her in 2 lists and they/them in 1 list.

Participants read descriptions of each character in the frame *[Name] uses [pronouns]. [Name] works as a [job] and has a [pet].* After a brief distractor task consisting of simple math questions, there were two tasks measuring learning of pronouns: In the memory task, participants were given the character's name and answered multiple-choice questions about their pronouns, job, and pet (e.g*., What pronouns does Emily use? What is Emily's job? What kind of pet does Emily have?*). The 3 questions about each character always appeared together, the order of the 12 characters was pseudo-randomized, and the orders of the choices for each question were randomized. In the production task, participants saw sentence fragments for each character (in a randomized order) that included their name and job (e.g., *After Emily got home from working as a teacher…*). Participants were asked to complete the sentence in a way that made sense to them, to measure which pronouns (if any) they used to refer to the character (e.g., …*they made dinner*). Finally, participants were shown 3 characters, 1 for each pronoun, and asked what they would say if they were introducing the character to someone. This task measured if participants used a pronoun in reference to the character or specifically mentioned which pronouns the character used, as a pilot for more open-ended prompts.

1.2.2 Predictions

Given that people who use they/them pronouns report high rates of both unintentional errors and intentional misgendering (Cordoba, 2020; Goldberg et al., 2019; James et al., 2016; McLemore, 2018; The Trevor Project, 2020), we predict that he/him and she/her will be remembered and produced more accurately than they/them. This outcome could be observed for one or more reasons: Participants may be unfamiliar with singular *they*, or familiar with comprehending it but unused to producing it. Singular *they* is also less frequent than *he* and *she*,

and as a result may be more difficult to use, even for speakers already familiar with it. Additionally, if participants avoid using they/them as an option, instead choosing the pronouns typically associated with the character's name, accuracy for they/them would also be lower than accuracy for he/him and she/her. Alternatively, the relative novelty of they/them may improve memory, as distinctive information tends to be remembered better (von Restorff, 1933; Wallace, 1965). Under this account, accuracy remembering and producing they/them would be higher than for he/him and she/her.

We hypothesize that learning to use singular *they* requires a change from inferring a person's pronoun (*he* or *she*) based on semantic/conceptual features of a person (e.g., Antón-Méndez, 2010) or based on morphosyntactic gender associated with a person's name (e.g., Schmitt et al., 1999), and instead recalling episodic information about a person's stated pronouns. In the context of this experiment, the gender association of the character's name cannot predict whether that character uses *he/she* or *they*, meaning that the only way to consistently produce the correct pronouns is to remember the information from the character introductions. As a result, we predict that correctly remembering that a character uses they/them in the multiple-choice task should predict correctly producing *they* in the sentence completion task. Alternatively, pronoun choice in the production task may not be influenced by episodic memory for which pronouns a character uses. This would occur if, in language production, a speaker attempts to infer the character's pronouns based on their name rather than retrieving them from memory, or if a speaker chooses to not produce singular *they*. In this scenario, accuracy in the memory task would not predict accuracy in the production task.

1.2.3 Results

Responses were analyzed using logistic mixed-effects regression models using *lme4* in R (Bates et al., 2015; R Core Team, 2023), with the first model analyzing memory accuracy (Table 1.1), the second analyzing production accuracy (Table 1.2), and the third analyzing a relation

between memory and production accuracy (Table 1.3). The fixed effect of pronoun was coded with orthogonal Helmert contrasts; the first contrast compared they/them to he/him and she/her, and the second contrast compared he/him to she/her. In the model that included memory accuracy as a fixed effect, it was mean-center effects coded. The maximal models included by-participant and by-item (defined as the 12 names) random intercepts and slopes for pronoun (Baayen et al., 2008; Barr et al., 2013). The *buildmer* package (Voeten, 2023) was used to select the maximal converging models; all three final models included by-participant intercepts, the memory and production models also included by-participant slopes, and the memory model also included by-item intercepts.

Memory. For accuracy in the multiple-choice memory task (Table 1.1), participants responded more accurately than inaccurately across all three pronoun conditions ($\beta = 0.77$, $z = 7.36$, $p < .001$). He/him and she/her ($M = 0.77$) were remembered significantly more accurately than they/them ($M = 0.44$) ($\beta = 1.64$, $z = 9.54$, $p < .001$). Errors were asymmetric: not remembering that a character used they/them was more common than incorrectly attributing they/them to a character (Figure 1.1). Recall that just as each character was associated with one of three pronouns, each character was also associated with one of three pets (Figure A1.1). For they/them characters, pronoun accuracy was compared to pet accuracy ($M = 0.39$)—which exhibits similar distributional characteristics—showing no significant difference ($\beta = 0.21$, $z = 1.35$, $p = .18$). Accuracy for the characters' 12 possible jobs ($M = 0.21$) was not at floor, suggesting that overall, the task was not too difficult for participants. The pet and job questions are discussed in more detail in the appendix (§A.1.1).

Table 1.1: Experiment 1A: Memory. Model results for the effect of Pronoun on Memory Accuracy.

Experiment 1A: Memory

	Memory Accuracy			
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	0.768	0.104	7.359	**<0.001**
Pronoun: They (-.66) **vs He** (+.33) **+ She** (+.33)	1.638	0.172	9.544	**<0.001**
Pronoun: He (-.5) vs She (+.5)	0.129	0.217	0.596	0.551
Random Effects				
$T_{00\ Participant}$	0.443			
$T_{00\ Name}$	0.008			
$T_{11\ Pronoun\ (They\ vs\ He\ +\ She)\ |\ Participant}$	0.634			
$T_{11\ Pronoun\ (He\ vs\ She)\ |\ Participant}$	0.193			
$\rho_{01\ Pronoun\ (They\ vs\ He\ +\ She)\ |\ Participant}$	-0.425			
$\rho_{01\ Pronoun\ (He\ vs\ She)\ |\ Participant}$	0.122			
$N_{\ Participant}$	102			
$N_{\ Name}$	12			
Observations	1224			

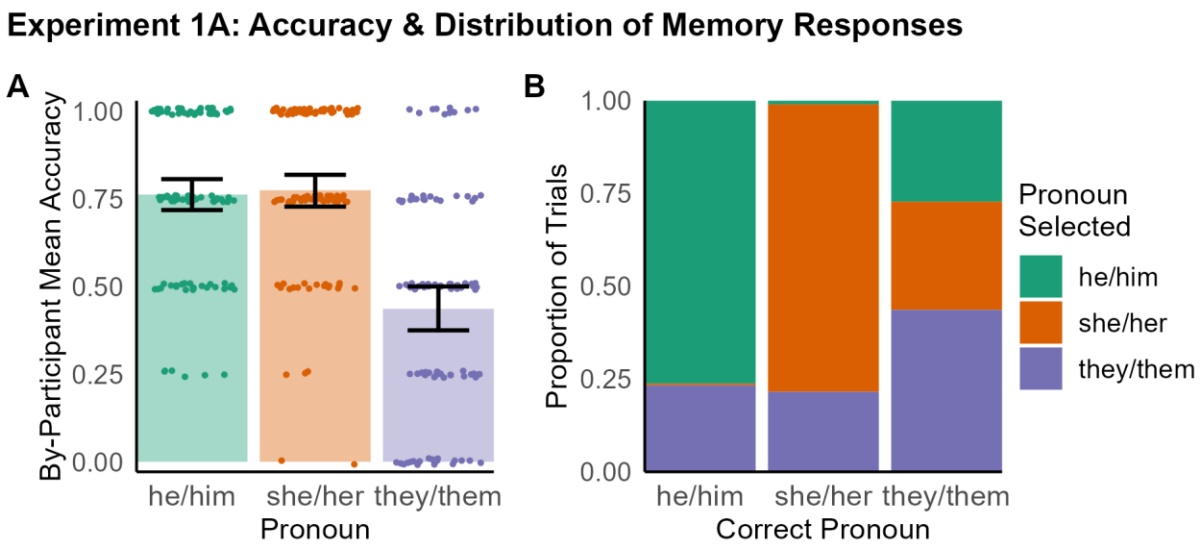


Figure 1.1: Experiment 1A: Memory. [A] Pronoun accuracy in the multiple-choice memory task. By-participant means are shown as points; error bars indicate 95% CIs calculated over the by-participant means. [B] Distribution of memory responses, with the correct pronoun on the X axis and the selected pronoun by color.

Production. Responses were coded by whether the sentence continuation used he/him, she/her, they/them, or no pronouns to refer to the character (Figure 1.2). Responses that did not include a pronoun referring to the character either used the character's name as the subject of the continuing clause (e.g., *After Emily got home from working as a teacher…Emily made dinner*), or were ungrammatical continuations with no subject (e.g., *…made dinner*). Because these responses were infrequent (2% of trials) and evenly distributed between pronoun conditions, they are included in the analysis (Table 1.2) as incorrect responses. Participants responded more accurately than inaccurately across all three pronoun conditions ($\beta = 1.33$, $z = 6.73$, $p < .001$). He/him and she/her ($M = 0.84$) were produced significantly more accurately than they/them ($M = 0.29$) ($\beta = 4.14$, $z = 8.80$, $p < .001$). They/them was incorrectly produced as he/him or she/her each about one third of the time, while he/him and she/her were each incorrectly produced as they/them about an eighth of the time. Approximately 60% of participants produced singular *they* at least once, regardless of accuracy.

Table 1.2: Experiment 1A: Production. Model results for the effect of Pronoun on Production Accuracy.

Experiment 1A: Production

	Production Accuracy			
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	1.330	0.198	6.726	**<0.001**
Pronoun: They (-.66) **vs He** (+.33) **+ She** (+.33)	4.142	0.471	8.802	**<0.001**
Pronoun: He (-.5) vs She (+.5)	-0.157	0.472	-0.333	0.739
Random Effects				
$\tau_{00\ Participant}$	1.006			
$\tau_{11\ Pronoun\ (They\ vs\ He\ +\ She)\ |\ Participant}$	12.336			
$\tau_{11\ Pronoun\ (He\ vs\ She)\ |\ Participant}$	0.317			
$\rho_{01\ Pronoun\ (They\ vs\ He\ +\ She)\ |\ Participant}$	0.657			
$\rho_{01\ Pronoun\ (He\ vs\ She)\ |\ Participant}$	-0.945			
$N_{\ Participant}$	102			
Observations	1224			

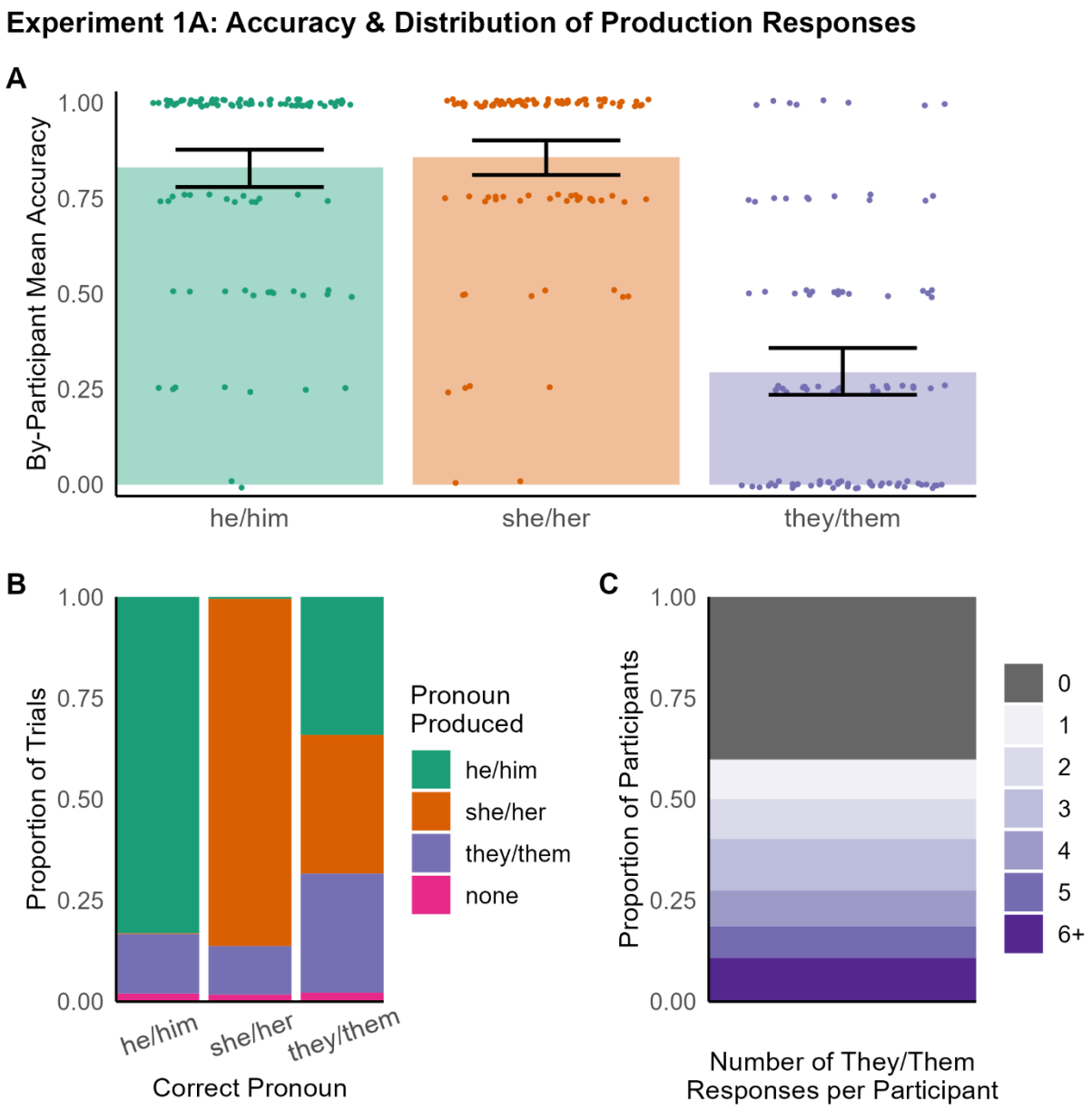


Figure 1.2: Experiment 1A: Production. [A] Pronoun accuracy in the written sentence completion task. By-participant means are shown as points; error bars indicate 95% CIs calculated over the by-participant means. [B] Distribution of pronoun production responses, with the correct pronoun on the X axis and the produced pronoun by color. [C] Number of times each participant produced singular they *(correct = 4), regardless of accuracy.*

Memory Predicting Production. Extending the second model to test the effects of pronoun and memory accuracy on production accuracy (Table 1.3) showed that participants were more likely to produce the correct pronoun when they had correctly remembered it (β = 1.25, z = 7.40, p < .001). Memory accuracy interacted with pronoun for the comparison between they/them and he/him + she/her (β = -0.82, z = -2.44, p < .05), such that memory improved production more for they/them characters than for he/him + she/her characters (Figure 1.3). Remembering but not producing they/them was more common than producing it but not remembering it.

Table 1.3: Experiment 1A: Memory Predicting Production. Model results for the effects of Pronoun and Memory Accuracy on Production Accuracy.

Experiment 1A: Memory Predicting Production

	Production Accuracy			
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	0.751	0.102	7.331	**<0.001**
Pronoun: They (-.66) **vs He** (+.33) **+ She** (+.33)	2.607	0.176	14.781	**<0.001**
Pronoun: He (-.5) vs She (+.5)	0.223	0.212	1.054	0.292
Memory Accuracy: Wrong (-.5) **vs Right** (+.5)	1.245	0.168	7.405	**<0.001**
Pronoun (They vs He + She) * Memory Accuracy	-0.820	0.336	-2.436	**0.015**
Pronoun (He vs She) * Memory Accuracy	-0.064	0.429	-0.149	0.881
Random Effects				
$T_{00\ Participant}$	0.347			
$N_{\ Participant}$	102			
Observations	1224			

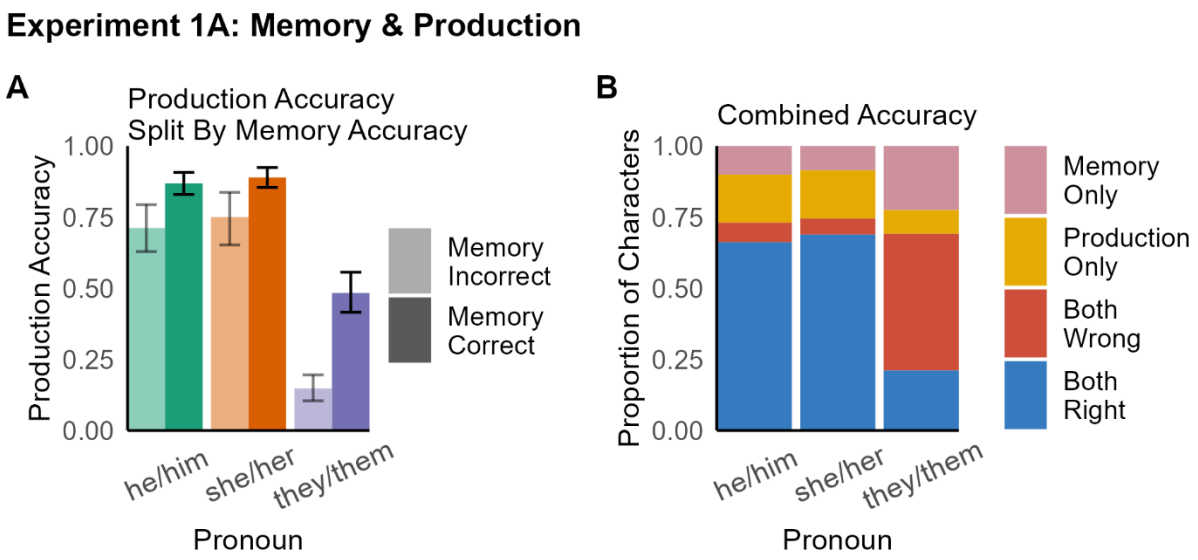


Figure 1.3: Experiment 1A: Memory Predicting Production. [A] Production accuracy split by memory accuracy. Error bars indicate 95% CIs calculated over trials. [B] Distribution of combined memory and production accuracy.

1.3 Experiment 1B

In the first experiment, people learned about a set of 12 characters whose pronouns—they/them or the expected he/him or she/her—could not be predicted from their name or other cues. These results demonstrate that, after only a brief exposure, people can learn that a character uses they/them pronouns and retrieve this information from memory to produce singular *they* in reference to them. Remembering that the character used they/them was a strong predictor of accurate pronoun production, with 48% of production trials correct when the memory trial had been correct, but only 15% of production trials correct when the memory trial had been incorrect. These results are congruent with a model where learning to use singular *they* requires a change from inferring a person's pronouns (*he* or *she*) based on semantic/conceptual features of a person or based on morphosyntactic gender associated with a person's name, and instead recalling episodic information about a person's stated pronouns.

While memory accuracy did predict production accuracy, it was not a guarantee. For they/them characters, memory was significantly more accurate than production (β = -0.80, z = -4.77, p < .001). Conversely, for he/him and she/her characters, memory was significantly less accurate than production (β = 0.56, z = 4.19, p < .001) (§A.1.1). I interpret these findings as demonstrating that memory for which pronouns a person uses and accuracy when producing a person's pronouns are distinct but interacting processes. However, another consideration is that the order of the memory and production tasks is in play. Consider that participants only had one brief exposure to information about the characters, then solved math problems for around five minutes, then answered 36 multiple-choice questions about each character's job, pet, and pronouns. By the time participants completed the production task, they may have begun to forget which characters used they/them pronouns, instead relying more on inferences from the characters' names. This would also result in higher production accuracy for he/him + she/her and lower production accuracy for they/them, but not necessarily due to different task demands between remembering which pronoun the character uses and selecting the correct pronoun to produce. A replication study that counterbalanced task order was conducted to ensure that the pattern of results in Experiment 1A was not determined by task order. While participants in Experiment 1A completed the memory task before the production task, participants in Experiment 1B completed the production task before the memory task. In all other respects, Experiment 1B was identical.

1.3.1 Methods

101 Vanderbilt undergraduates completed the study for partial course credit (M_{age} = 19.28, SD_{age} = 1.18). 72 were women, and 29 were men, as reported in a free response box. 82 rated their English ability as "native (learned from birth)," 17 as "fully competent…but not native," and 2 as "limited but adequate competence." The only difference from Experiment 1A was switching the order of the memory and production tasks.

1.3.2 Results

Responses were analyzed using the same logistic mixed-effects model software and specifications as in Experiment 1A, although the maximal random effects structures that converged differed slightly in this data set. Figure 1.4 highlights results for the they/them characters; results for all three pronouns, parallel to the figures in Experiment 1A, are shown in the appendix (§A.1.3). To compare these results to those of Experiment 1A, an additional trio of models with the combined data included Experiment as a mean-centered fixed effect.

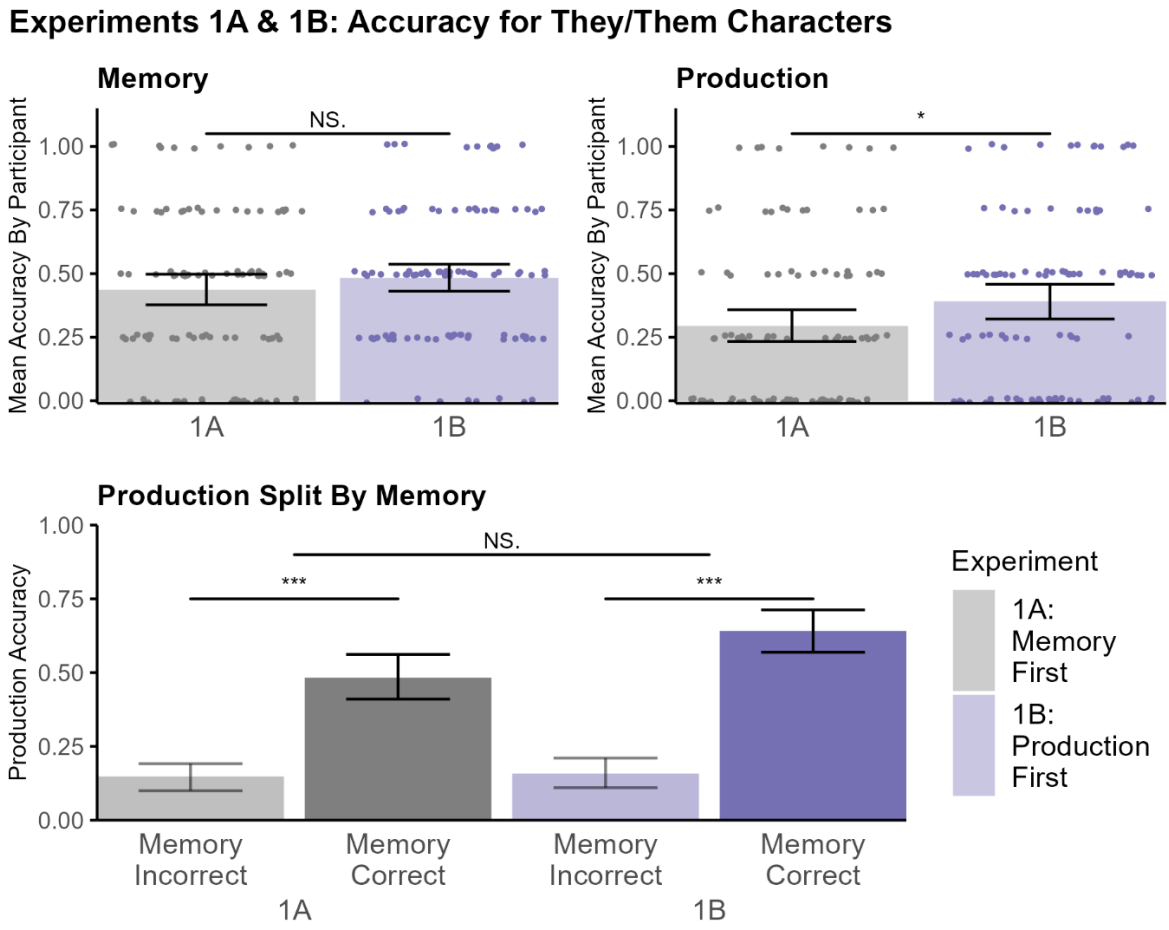


Figure 1.4: Experiments 1A & 1B: Memory and Production. [A] By-participant mean memory accuracy for they/them characters, split by task order. [B] By-participant mean production accuracy for they/them characters, split by task order. [C] Production accuracy for they/them characters, split by memory accuracy and task order. Error bars indicate 95% CIs.

Memory. Analyzing the effects of pronoun condition on accuracy in the multiple-choice memory task (Table A1.3), participants responded more accurately than inaccurately across all three pronouns (β = 0.95, z = 10.03, p < .001). He/him and she/her (M = 0.79) were remembered more accurately than they/them (M = 0.48) (β = 1.55, z = 10.88, p < .001). Comparing accuracy for the 3 pets, which was designed as a control for pronouns, indicated that for he/him + she/her characters, pronoun accuracy was higher than pet accuracy (β = 1.74, z = 13.70, p < .001). For they/them characters, the difference between pronoun and pet accuracy was not significant (β = 0.28, z = 1.85, p = .06). Memory for the characters' 12 jobs (M = 0.29, SD = 0.45) was again above floor and was numerically higher than Experiment 1A (Figure A1.2). When comparing memory accuracy between experiments (Table A1.7), neither the main effect of experiment (β = 0.18, z = 1.34, p = .18) nor its interactions with pronoun (β = -0.04, z = -0.22, p = .83) were significant. Overall, completing the memory task last showed little effect on accuracy.

Production. When analyzing the effects of pronoun on accuracy in the sentence completion task (Table A1.4), responses that did not include a pronoun referring to the character were again infrequent (3%) and are included in the analysis as incorrect responses. Participants used the correct pronoun to refer to the character more often than not across pronoun conditions (β = 1.11, z = 8.97, p < .001). He/him and she/her (M = 0.84) were produced more accurately than they/them (M = 0.39) (β = 2.48, z = 14.79, p < .001). Overall accuracy was not significantly different from the main experiment (β = 0.13, z = 1.29, p = .20). However, the interaction between pronoun and experiment was significant (β = -0.45, z = -2.22, p < .05), such that the difference in accuracy between they/them and he/him + she/her was reduced when the production task came first, compared to when the production task came second (Table A1.8).

Memory Predicting Production. Testing the effects of pronoun condition and memory accuracy on production accuracy (Table A1.5) showed that, in addition to the effects described above, participants were more likely to produce the correct pronoun if they also remembered it (β = 1.49, z = 9.27, p < .001). Memory accuracy interacted with pronoun for the comparison

between they/them and he/him + she/her (β = -1.16, z = -3.58, p < .001), such that memory improved production more for they/them characters than for he/him + she/her characters. Remembering but not producing they/them was again more common than producing but not remembering it. Comparing between experiments (Table A1.9), memory accuracy did not interact with task order (β = 0.27, z = 1.21, p = .23).

Figure 1.5 shows each participant's mean difference in accuracy between the two tasks. For they/them characters, participants in both experiments were more accurate in the memory task. Conversely, for he/him + she/her characters, participants in both experiments were more accurate in the production task. Both patterns are consistent with participants forgetting that a character uses they/them or choosing not to use singular *they*, then defaulting to the pronouns associated with the character's name.

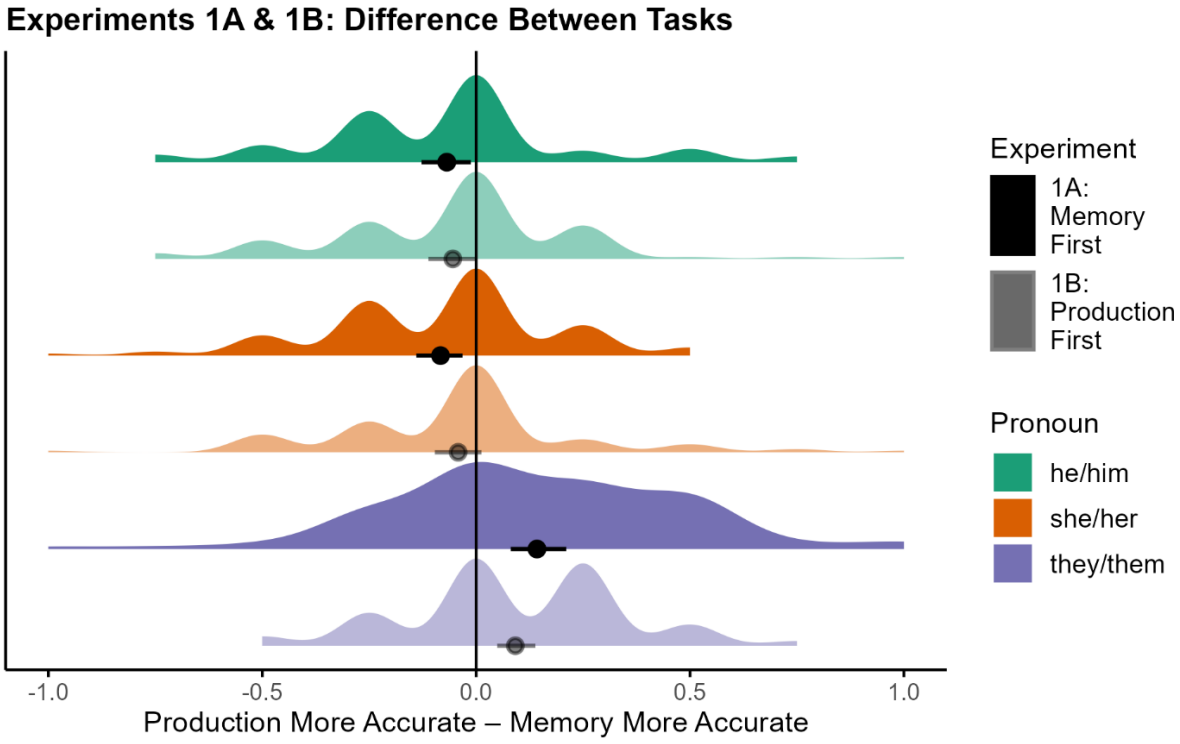


Figure 1.5: Experiments 1A & 1B: Task Differences. Differences between memory accuracy and production accuracy for each participant, split by character pronoun and task order.

In a linear regression with pronoun and task order predicting the by-participant mean task differences (Table A1.10), there were no significant effects of task order (β = 0.00, z = 0.07, p = .97) or its interaction with pronoun (β = 0.08, z = 1.55, p = .12).

Reliability. To estimate the internal reliability of the memory and production tasks, I used the Bayesian mixed-effects model approach described in Staub (2021). The trials from Experiments 1A & 1B were split in half, so that each half included 2 he/him, 2 she/her, and 2 they/them characters for each participant. Pronoun was coded as 2 variables: the first comparing they/them (-.66) to he/him (+.33) and she/her (+.33) in even trials, with odd trials coded as 0, and the second comparing they/them to he/him + she/her in odd trials, with even trials coded as 0. The *brms* package (Bürkner, 2017) fit 4 Bayesian mixed-effects models: memory and production accuracy for 1A and 1B. Each model included the odd and even trial pronoun variables as fixed effects and by-participant slopes. All models were fit using the default priors and 4 chains, each chain with 4000 iterations, of which 2000 were warm-up. The random slope estimates represent the relative accuracy of they/them compared to he/him + she/her for each participant (Figure A1.5). For the memory task, the correlation between slopes was low in both 1A, r = 0.34 [-0.65, 0.92], and 1B, r = 0.45 [-0.49, 0.95], indicating poor internal reliability. For the production task, the correlation between slopes was high in 1A, r = 0.91 [0.77, 0.99], but medium in 1B, r = 0.64 [0.34, 0.88]. Internal reliabilities below 0.80 for the memory task in both experiments and for the production task in 1B indicate that analyzing individual differences would not be warranted.

1.4 Discussion

Experiment 1 investigated how people associate pronouns with a person, in a context where producing the correct pronoun required recalling stated information about the character's pronouns, instead of choosing the pronoun based on the gender association of the name or an inference about the gender of the character. Participants learned about a set of 12 characters whose pronouns—they/them or the expected he/him or she/her—could not be predicted from their

name or other cues. Memory for pronouns was measured in a multiple-choice task, and production of pronouns was measured in a written sentence completion task. After only a brief exposure to the characters, participants correctly recalled that the character used they/them pronouns in about half of trials and correctly produced singular *they* in reference to the character in about a third of trials. Participants originally completed the memory task before the production task, and a follow-up experiment presented the production task first, to rule out the possibility that differences between memory and production accuracy were caused by task order. In both experiments, remembering that the character used they/them pronouns strongly predicted—but did not guarantee—producing singular *they*. These results provide support for a model where speakers can select pronouns by retrieving information from episodic memory about a person's stated pronouns, instead of selecting pronouns based on morphosyntactic gender information from the name or a gender inference about the person.

One limitation is that Experiment 1 was conducted with undergraduate students at a private university, which represents a particular subset of English speakers located in the U.S.: primarily ages 18–22 and highly educated. Prior data allows us to infer that the participants are more likely to be socially liberal, have prior knowledge about LGBTQ+ topics, and to consider singular *they* acceptable (Minkin & Brown, 2021; Parker et al., 2019). However, more extensive demographic and language experience data were not included in the first experiment, because investigation of individual differences would first need to establish the internal reliability of the measures, in order to then be able to associate them with other person-specific variables.

There is an increasing awareness in psycholinguistics, and in cognitive psychology more broadly, of the importance of demonstrating sufficient internal reliability as a first step in conducting individual differences research (Cronbach, 1957; Hedge et al., 2017; Staub, 2021). This often proves difficult, because reliably measuring effects of experimental manipulations and reliably measuring individual differences are mathematically at odds with each other. Individual differences measures aim to characterize *between*-participant variability (e.g., the extent to which

someone endorses the gender binary and gender essentialism), and reliability means that the measure ranks participants consistently. This requires a low degree of *within*-participant variability: participants' responses would be consistent both within trials in the same experiment and if they completed the experiment again at a later date. Conversely, most cognitive psychology measures aim to characterize changes in behavior in different experimental contexts (e.g., singular *they* takes more time to read than plural *they*), which is *within*-participant variability. In this context, reliability means that the effect replicates in different experiments. This requires a low degree of *between*-participant variability, where most participants show the same pattern of responses. Hedge et al. (2017) describe this incompatibility as the "reliability paradox," because "robust experimental paradigms…are likely to be sub-optimal for correlational studies for *the same reasons* that they produce robust experimental effects."

If a measure shows low internal reliability—where each participant's responses are not consistent within halves of the same experiment, or when completing the experiment twice—testing for individual differences would be possible, but not warranted (Hedge et al., 2017). One of the primary issues is that the reliability of the measures constrains the maximum expected true correlation that one would expect to find (Spearman, 1904), meaning that correlations with a low-reliability task will be underestimated. This is the position we find ourselves in with Experiment 1. Analyzing the data using Staub (2021)'s split-half mixed-effects model approach showed that the internal reliability of the memory task ($r_A = 0.34$, $r_B = 0.45$) was too low to warrant individual differences analysis, and the reliability of the production task was high enough in Experiment 1A but not 1B ($r_A = 0.91$, $r_B = 0.64$). While increasing the number of trials per participant can increase reliability (Liceralde & Watson, 2023), this is not feasible with a memory task, as adding more characters to learn about would make the experiment too difficult for participants. Overall, this means that while people clearly do vary in their accuracy remembering and producing singular *they*, it is not clear how much of this variability arises from theoretically-interesting factors (e.g., language and gender beliefs), how much arises from general psychological factors (e.g., ability in

memory tests, attention to the experiment), and how much is random noise. Instead of pursuing individual differences questions with the character learning task, we turn to investigating how presenting information about singular *they* can support accurate memory and production.

CHAPTER 2

Experiment 2: Effects of a PSA and Usage Modeling on Memory and Written Production

2.1 Motivation

The results of Experiment 1 suggest that people can learn to associate pronouns with a person, but that accuracy for they/them remains lower than for he/him and she/her. Although remembering which characters used they/them was a strong predictor of producing singular *they,* accuracy in the sentence completion task was significantly lower than in the multiple-choice memory task. Experiment 2 investigated what kinds of exposure can support accurately remembering and producing singular *they*. The first factor tested is the role of conceptual knowledge about singular *they* and discussing gendered language preferences. Recent results show that participants are more likely to interpret *they* as the intended singular, instead of plural, after being told explicitly that the character uses they/them pronouns (Arnold et al., 2021, described in §0.4.4). This is also supported by prior experiments about the generic masculine: When a course instructor included information about why they would be using generic *she* instead of generic *he* (Adamsky, 1981), and when alternatives were taught as options to students (Flanagan & Todd-Mancillas, 1982), students were less likely to use generic *he* in their assignments and more likely to use gender-neutral alternatives or generic *she*. Similarly, in German (where nouns are gender-marked) reading brief arguments in favor of gender-neutral language increased participants' use of gender-neutral generic nouns (Koeser & Sczesny, 2014).

The second factor tested is exposure. As singular *they* becomes more common and accepted (Balhorn, 2004; Camilliere et al., 2021; Hekanaho, 2020; Minkin & Brown, 2021; Parker et al., 2019), speakers are increasingly likely to be exposed to it via media and social circles, and many of these instances do not come prefaced with a discussion about pronouns or gender

identity. Potentially-comparable results from studies about non-sexist language reforms are mixed: students who saw alternatives to generic masculine forms modeled in task instructions increased their use of non-sexist forms, but did not decrease their use of generic masculine forms (Cronin & Jreisat, 1995). In German, women were more likely to use alternatives to generic masculine role nouns after reading a text modeling them, but men did not change their language use until the instructions drew their attention to the gendered language used (Koeser et al., 2015).

2.2 Methods

The design and analysis plan were preregistered on the Open Science Framework (10.17605/OSF.IO/3DZE4). Materials, de-identified data, and analysis code are available at this dissertation's Github repository.

2.2.1 Participants

427 responses were collected from Amazon Mechanical Turk, completing a task that took approximately 20 minutes. Participants were required to be in the U.S. and comfortable reading and writing in English. 107 participants were excluded for nonsensical responses in the sentence completion task, for a total of 320 participants in the final data set. As in Experiment 1, participants were asked about their age (M_{age} = 38.79, SD_{age} = 11.54), gender (194 male, 125 female, 1 did not provide), and English experience (304 "native (learned from birth)," 14 "fully competent…but not native," 1 "limited but adequate competence," 1 "some familiarity") in order to characterize the sample (§B.1.1).

2.2.2 Materials & Procedure

Participants read 1 of 2 500-word PSAs (§B.2). The pronoun PSA was modified from a GLSEN resource and discussed talking about gendered language preferences, using singular *they*, and responding to misgendering someone (GLSEN, 2020). The neutral PSA was modified

from a Humane Society resource and discussed the importance of spaying/neutering cats and dogs (Humane Society, 2020). Participants also read 2 fictional biographies, which made repeated third-person reference to a single character, in order to model pronoun use without explicitly commenting on it (§B.3). The character in the first biography had a feminine name and was referred to with they/them or she/her pronouns (4 subject, 1 object, 4 possessive). The character in the second biography had a masculine name and was referred to with they/them or he/him pronouns (7 subject, 7 possessive). The other materials were identical to Experiment 1. Participants read 1 PSA and 1 pair of biographies (2 they/them characters, or 1 he/him character and 1 she/her character). These were crossed to create 4 between-participants conditions [**PSA: Gendered Language vs Unrelated; Biographies: They vs He/She**]. Participants then completed the same pronoun memory and production tasks as in Experiment 1. Participants were randomly assigned to 1 of 3 lists within each condition, counterbalancing the name-pronoun combinations. The experiment was coded and hosted using PCIbex (Zehr & Schwartz, 2018).

2.3 Predictions

The PSA contains information about why paying attention to gendered language matters, mentions singular *they* as an option and shows examples of its usage, and provides scripts for talking about gendered language preferences (GLSEN, 2020). This addresses conceptual knowledge about singular *they* and misgendering, chosen to be similar to Diversity, Equity, and Inclusion materials that people may see in their schools or workplaces. If learning or being reminded of this information affects language use, we predict that participants who read the gendered language PSA will be more accurate at remembering and producing they/them, compared to participants who read the unrelated PSA.

As singular *they* becomes more common and accepted (Balhorn, 2004; Camilliere et al., 2021; Hekanaho, 2020; Minkin & Brown, 2021; Parker et al., 2019), speakers are increasingly likely to be exposed to it via media and social circles, and many of these instances do not come

prefaced with a discussion about gendered language or gender identity. As such, the biographies model the use of singular *they*, but do not explicitly call attention to it. The biography genre allows for repeated reference to one individual, giving participants multiple examples and making it more straightforward to interpret *they* as singular and not plural. If seeing singular *they* modeled supports learning, we predict that participants who read the stories about characters referred to with they/them pronouns will be more accurate at remembering and producing singular *they*, compared to participants who read the stories about characters referred to with he/him and she/her pronouns.

2.4 Results

Three logistic mixed-effects models analyzed Pronoun, PSA, and Biography predicting memory accuracy (Table 2.1), production accuracy (Table 2.2), and a model relating the two measures (Table A2.2). The fixed effects of PSA and Biography were mean-center effects coded; all other model specifications followed Experiment 1 (Baayen et al., 2008; Bates et al., 2015; R Core Team, 2023; Voeten, 2023). For all three models, the most complex random effects structure that converged included only by-item intercepts, and no by-participant effects.

2.4.1 Memory

In the multiple-choice memory task (Table 2.1), participants responded more accurately than not across pronouns and training conditions (β = 1.00, z = 21.47, $p < .001$). He/him and she/her (M = 0.77–0.82 between PSA and Biography conditions) were remembered more accurately than they/them (M = 0.51–0.59) (β = 1.22, z = 16.02, $p < .001$). There was no difference in accuracy between he/him and she/her, and they/them was misremembered as he/him and she/her at similar rates (Figure 2.1A). The lower accuracy of they/them compared to she/her and he/him (Figure 2.1B) was attenuated when participants read the gendered language PSA compared to the unrelated PSA (β = -0.36, z = -2.38, $p < .05$).

Table 2.1: Experiment 2: Memory. Model results for the effects of Pronoun, PSA, and Biography on Memory Accuracy.

Experiment 2: Memory

	Memory Accuracy			
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	0.996	0.046	21.467	**<0.001**
Pronoun: They (-.66) **vs He** (+.33) **+ She** (+.33)	1.218	0.076	16.023	**<0.001**
Pronoun: He (-.5) vs She (+.5)	0.140	0.111	1.265	0.206
Biography: He/She (-.5) **vs Biography: They** (+.5)	-0.176	0.076	-2.304	**0.021**
PSA: Unrelated (-.5) vs PSA: Gendered Language (+.5)	0.022	0.076	0.289	0.772
Pronoun (They vs He + She) * PSA	-0.362	0.152	-2.381	**0.017**
Pronoun (He vs She) * PSA	0.121	0.199	0.609	0.543
Pronoun (They vs He + She) * Biography	-0.161	0.152	-1.060	0.289
Pronoun (He vs She) * Biography	0.049	0.199	0.245	0.807
PSA * Biography	-0.073	0.153	-0.481	0.630
Pronoun (They vs He + She) * PSA * Biography	-0.131	0.304	-0.431	0.666
Pronoun (He vs She) * PSA * Biography	-0.088	0.398	-0.220	0.826
Random Effects				
$\tau_{00\ Name}$	0.008			
N_{Name}	12			
Observations	3840			

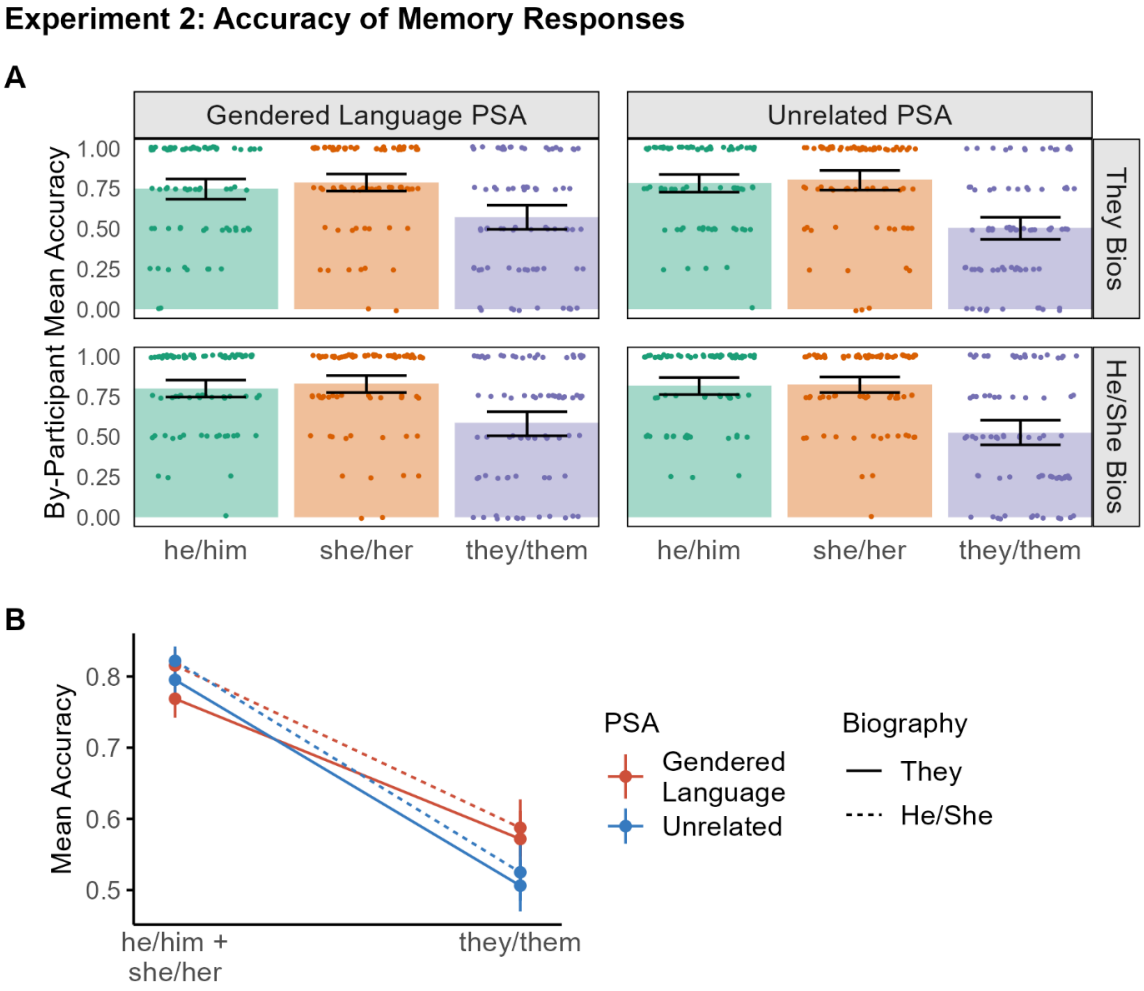


Figure 2.1: Experiment 2: Memory. [A] Pronoun accuracy in the multiple-choice memory task, split by PSA and Biography conditions. By-participant means are shown as points; error bars indicate 95% CIs calculated over the by-participant means. [B] Means and 95% CIs of memory accuracy for he/him + she/her characters and they/them characters, comparing PSA and Biography conditions. The distribution of responses is shown in the appendix (Figure A2.2).

Participants also learned that each character had 1 of 3 pets, which was designed to have the same distributional characteristics but be less marked in comparison to the 3 pronouns. As in Experiment 1, there was no significant difference between accuracy for they/them characters' pets (M = 0.54) and pronouns (β = 0.05, z = 0.48, p = .63). Accuracy for the 12 possible jobs was relatively high (M = 0.37), confirming that the experiment was not too difficult for participants. Job and pet accuracy are discussed in more detail in the appendix (§A.2.1).

2.4.2 Production

Responses were coded by whether the sentence continuation used he/him, she/her, they/them, or no pronouns to refer to the character (Figure 2.2). Responses that did not include a pronoun were 4% of the data and are included in the analysis as incorrect responses (Table 2.2). Across all conditions, participants produced the correct pronoun more often than not (β = 0.70, z = 13.47, p < .001). He/him and she/her (M = 0.79–0.90 between PSA and Biography conditions) were produced more accurately than they/them (M = 0.11–0.33) (β = 3.16, z = 32.95, p < .001). He/him was produced somewhat more accurately than she/her (β = -0.26, z = -2.11, p < .05). The relative difficulty of they/them was attenuated with the gendered language PSA (β = -1.91, z = -10.00, p < .001), and there was a significant interaction between PSA and Biography (β = 0.43, z = 2.35, p < .05). These effects were qualified by a three-way interaction between Pronoun, PSA, and Biography (β = 0.88, z = 2.31, p < .05). A follow-up analysis probing this interaction found that the gendered language PSA reduced the relative difficulty of they/them more when paired with the biographies that used he/him and she/her (β = -2.35, z = -8.50, p < .001) than when paired with the biographies that used they/them (β = -1.47, z = -5.57, p < .001). However, examining the means for the two conditions with the gendered language PSA (red in Figure 2.2B) indicates that the difference in relative accuracy for they/them compared to he/him + she/her is due to Biography affecting accuracy for he/him + she/her characters, but not accuracy for they/them characters. Finally, an exploratory analysis measured the proportion of participants who produced singular *they* at all, regardless of accuracy. Participants who read the gendered language PSA were more likely to produce singular *they* at least once (β = 0.93, z = 3.95, p < .001), with proportions rising from 26% and 35% in conditions that read the unrelated PSA to 48% and 57% in conditions that read the gendered language PSA (Table A2.3).

Table 2.2: Experiment 2: Production. Model results for the effects of Pronoun, PSA, and Biography on Production Accuracy.

Experiment 2: Production

	Production Accuracy			
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	0.695	0.052	13.469	**<0.001**
Pronoun: They (-.66) **vs He** (+.33) **+ She** (+.33)	3.156	0.096	32.951	**<0.001**
Pronoun: He (-.5) **vs She** (+.5)	-0.261	0.123	-2.113	**0.035**
PSA: Unrelated (-.5) vs PSA: Gendered Language (+.5)	0.107	0.091	1.178	0.239
Biography: He/She (-.5) vs Biography: They (+.5)	-0.059	0.091	-0.653	0.514
Pronoun (They vs He + She) * PSA	-1.907	0.191	-9.999	**<0.001**
Pronoun (He vs She) * PSA	0.266	0.227	1.170	0.242
PSA * Biography	0.426	0.182	2.345	**0.019**
Pronoun (They vs He + She) * Biography	-0.160	0.191	-0.840	0.401
Pronoun (He vs She) * Biography	0.085	0.227	0.372	0.710
Pronoun (They vs He + She) * PSA * Biography	0.882	0.381	2.313	**0.021**
Pronoun (He vs She) * PSA * Biography	0.134	0.454	0.296	0.768
Random Effects				
$T_{00\ Name}$	0.007			
N_{Name}	12			
Observations	3840			

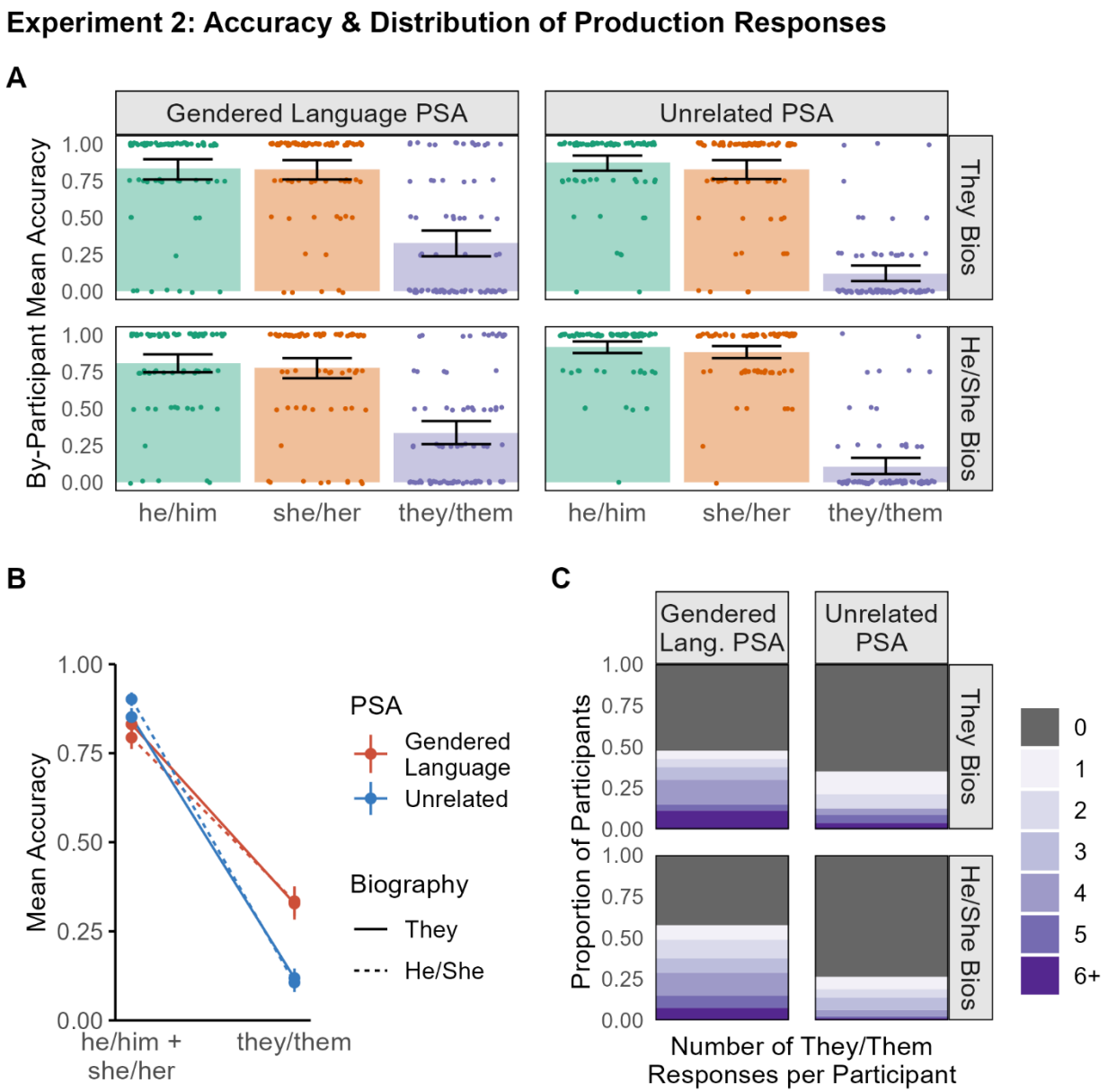


Figure 2.2: Experiment 2: Production. [A] Pronoun accuracy in the written sentence completion task, split by PSA and Biography conditions. By-participant means are shown as points; error bars indicate 95% CIs calculated over the by-participant means. [B] Mean production accuracy for he/him + she/her characters and they/them characters, split by PSA and Biography conditions. [C] Number of times each participant produced singular they, *split by PSA and Biography conditions. The distribution of all pronoun responses is shown in the appendix (Figure A2.2).*

2.4.3 Memory Predicting Production

The third model tested the effects of memory accuracy, pronoun, PSA, and Biography on production accuracy (Table A2.2). In addition to the effects described above, participants were more likely to accurately use a character's pronouns in the sentence completion task if they had remembered that character's pronouns in the multiple-choice task ($\beta = 0.83$, $z = 8.00$, $p < .001$). No other interactions with memory accuracy were significant. Examining the combined distribution of responses, it was again more common to remember but not produce they/them than to produce but not remember they/them (Figure 2.3).

2.5 Discussion

In Experiment 2, participants read either a PSA about gendered language or an unrelated topic, then two fictional biographies where both characters used they/them or one character used he/him and one character used she/her. Participants then completed the same character learning, memory, and production tasks as in Experiment 1. Reading the PSA about gendered language—which explained why people are talking more about their preferences for gendered language, how they/them pronouns work, and how to respond if someone corrects you—increased how likely participants were to produce singular *they* at least once and improved their accuracy when doing so. Seeing singular *they* modeled in the biographies did not directly affect memory or production, but did interact with the PSA. This demonstrates that while learning singular *they* may be difficult, it is not impossible, and even brief interventions can support this learning.

Compared to Experiment 1, which included undergraduates participating for course credit, Amazon MTurk participants vary more—particularly in terms of age, race, education, and socioeconomic status—but are still not fully representative of English speakers in the U.S. context (Arechar & Rand, 2021; Levay et al., 2016). MTurk participants lean more liberal than U.S. adults, and are more likely to agree that trans people are discriminated against and to support marriage equality and anti-discrimination laws for gay people (Chandler et al., 2019; Levay et al., 2016).

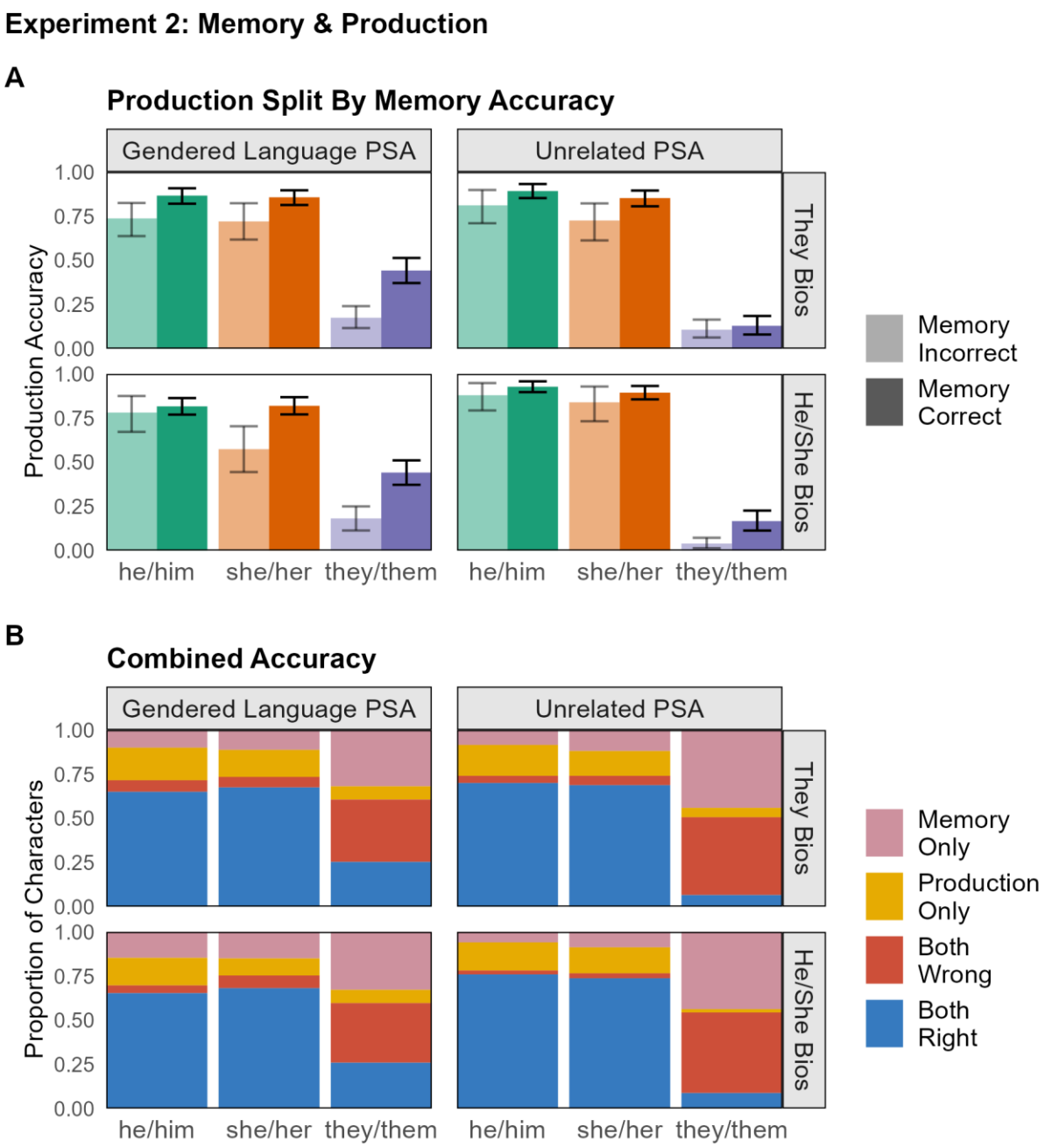


Figure 2.3: Experiment 2: Memory and Production. [A] Production accuracy, split by memory accuracy in the prior task, then by PSA and Biography conditions. The lighter colors indicate trials where memory had been incorrect, and the darker colors indicate trials where memory had been correct. Error bars indicate 95% CIs calculated over trials. [B] Distribution of combined memory and production accuracy, split by PSA and Biography conditions.

Most, but not all, participants in both experiments reported being native English speakers; while all participants in Experiment 1 were physically located in the U.S., in Experiment 2 the web-based restrictions that limited participation to U.S.-based individuals may not have been foolproof.

However, overall performance was, broadly speaking, similar across the two studies despite the sampling differences: While participants in the Unrelated PSA + He/She Biographies condition—the condition in Experiment 2 most similar to Experiment 1—were less likely to correctly produce *they* than participants in Experiment 1 (M_{1A} = 0.29, M_{1B} = 0.39, M_2 = 0.11), this is unlikely to be due to overall lower accuracy or attention to the task. Looking at the memory questions unrelated to pronouns, participants in Experiment 2 were numerically more accurate than participants in Experiment 1 for both the characters' jobs (M_{1A} = 0.21, M_{1B} = 0.29, M_2 = 0.37) and pets (M_{1A} = 0.41, M_{1B} = 0.43, M_2 = 0.54). While participants in Experiment 2 were less likely to have experience with singular *they* than participants in Experiment 1, the PSA manipulation was intended to provide participants with some of the social context around gendered language that they may have been less familiar with. In sum, these findings show that providing people with brief information about how they/them pronouns work, why people use them, and why people choose to talk directly about the gendered language they prefer did support people's use of singular *they*. Whether or not this effect may vary depending on participants' prior knowledge and experience is an area for future research.

The finding that reading a brief PSA increased both the overall usage and accuracy of singular *they* is promising, given that a PSA is an easily implemented and not time-intensive tool. Nevertheless, in order to be useful in applied contexts, future research will need to investigate whether the effects of the gendered language PSA and other learning interventions persist past the duration of an experiment. Future work should also investigate whether the effects on written production extend to spoken production.

CHAPTER 3

Experiment 3: Effects of Including Pronouns on Nametags and in Introductions on Spoken Production

3.1 Motivation

Encouraging people to include their pronouns when introducing themselves and providing space for people to indicate their pronouns in display names, email signatures, and nametags are common recommendations for making environments inclusive of TGD people (Richards & Barker, 2013). Options to specify your pronouns are currently included in many social media platforms such as Instagram and LinkedIn, in institutional platforms such as Brightspace and Slack, and in tools such as Zoom and Github. Ideally, group norms of indicating pronouns makes disclosure less marked, which supports the individuals who need to explicitly state their pronouns in order to avoid being misgendered—which is the majority of people who use they/them pronouns, in the majority of contexts.

Some recent research has investigated the effects of pronoun-sharing practices. Both TGD and other LGBQ+ people evaluated a potential workplace more positively when a biography of a staff member included that she used she/her pronouns, when she/her would have been assumed (I. R. Johnson et al., 2021). This suggests that other people indicating their pronouns can act as an identity-safety cue when TGD and LGBQ+ people are forming initial appraisals of an environment, so they may be more likely to choose that environment and be more comfortable being out.[12] Additionally, directly explaining that a character used they/them supported people's

[12] *However, see McGonagill (2023), discussed in §0.7, for data about how nonbinary people felt that being out would negatively affect their job search and how this was borne out in experiments with resume response rates and hiring manager surveys. When an otherwise-identical resume included they/them pronouns, the applicant received less interest and was rated as less qualified.*

ability to correctly comprehend *they* as singular, not plural (Arnold et al., 2021, see §0.4.4).

An unanswered question, however, is whether pronoun-sharing practices affect people's language *production*—specifically, if they reduce the frequency of misgendering and support accurate use of singular *they*. While creating an environment where TGD people feel safe asking for the correct gendered language and where others understand the usage of singular *they* are both good outcomes, EDI practices need to affect allies' concrete behavior (e.g., Kattari et al., 2018), beyond their knowledge and attitudes.

In the first two experiments, I have argued for a model where learning to use they/them pronouns requires retrieving information about a person's pronouns from episodic memory, instead of inferring pronouns from morphosyntactic features of their name or from an inference about their gender. Experiment 1 demonstrated that when a character's pronouns cannot be inferred from their name, people can learn to associate pronouns with the character and use that information in language production, and Experiment 2 demonstrated that accurately producing they/them pronouns can be supported by providing participants with information about why paying attention to a person's pronouns is important. Experiment 3 moves from testing an intervention that may influence how participants pay attention to and attempt to recall information about a person's pronouns to comparing different ways of presenting information about pronouns. The current experiment investigates two practices for providing explicit information about what pronouns a person uses: stating pronouns when introducing someone, which makes this information highly salient at the beginning of the conversation, and including pronouns on someone's nametag/display name, which keeps this information accessible throughout the conversation. Ideally, pronoun-sharing practices will support production of singular *they*. However, pronoun-sharing practices may not consistently impact pronoun production—at all, or at least not to a degree that has real-world applicability. The frequency of seemingly-counterintuitive errors like "she uses they/them" shows that speakers may have information about a person's pronouns available, but still produce the incorrect pronoun.

3.2 Methods

The design and analysis plan were preregistered on the Open Science Framework (10.17605/OSF.IO/BT7YN). Sources and attributions for the images are located in the appendix (§B.5.2), and the edited images are available upon request. The de-identified data and analysis code are available at this dissertation's Github repository.

3.2.1 Participants

A simulation-based power analysis using the *simr* package in R (Green & MacLeod, 2016) estimated the number of participants required to detect a 2-way interaction between the condition manipulation and pronoun type with the same effect size (β = -1.91, OR = 0.15) as the production task in Experiment 2. This indicated that 156 participants, each completing 30 trials, would have .93 [.91, .94] power at α = .05 to detect the interaction.

Participants were recruited from Prolific (Peer et al., 2022) and required to be over the age of 18, be native or advanced speakers of English, and have a device with a microphone to record audio. The task took approximately 20 minutes. 162 participants are included in the analysis, with an additional 11 participants excluded for stopping the study before completing 25 test trials, having technical errors saving data, or having >5 test trial recordings that did not include a response to the task. Participant demographics are described in Table 3.1.

3.2.2 Materials

Characters. Participants were introduced to 3 characters, each of whom was associated with a set of pronouns (1 he/him, 1 she/her, 1 they/them), a name, an image, a pictured brother character, and a pictured sister character (Figure B1). The character images were selected from *The Gender Spectrum Collection*, a stock photo library created to provide a diverse range of images of transgender and nonbinary people (Drucker, 2019). The sibling images were selected from a free-use stock photo database (§B.5.2). Each image was edited to show 1 person from

the shoulders up with a white background. Nametags were shown in white text on a black bar along the bottom of the image, resembling display names in the Zoom interface. For characters, the nametags showed the first name and, depending on the condition, their pronouns. For siblings, the nametags showed *[character name]'s [brother/sister]*, in order to limit the number of names participants needed to learn and to elicit possessive pronouns referring to the character.

Across the lists, there were 6 images for the main characters. These were selected by conducting a norming study on 12 images (§B.5.1). Participants (N = 30) on Prolific saw each image paired with a sentence completion prompt, which referred to the pictured person as "this person" and did not include a name (e.g., *This person made a mug of tea. Before it cooled...*). No information about pronouns was included, either in the instructions or in the prompts. Participants wrote a completion to each prompt, in order to measure which pronouns, if any, were chosen to refer to the character (Table B1). Unlike in the first two studies, participants produced they/them frequently (71% of responses). This is potentially due to prompts not including names, as well as a different participant population on Prolific compared to MTurk. From the results of this norming study, 3 images where people did not produce she/her pronouns and 3 images where people did not produce he/him pronouns were selected. The goal was to create stimuli where we can expect that participants are primarily choosing between he/him and they/them or between she/her and they/them, not between all three, which may involve a different mechanism. This design allows us to test if including pronouns on nametags and in introductions increases accuracy for they/them, as compared to the pronouns that speakers would typically have defaulted to. This is the situation in which many people who use they/them find themselves, where they want people to stop using he/him or she/her pronouns. Additionally, the decision between he/him and she/her involves additional social dynamics that are outside the scope of the current study.

Participants were randomly assigned to 1 of 6 lists, in order to counterbalance the images and names associated with the character who uses they/them. Out of the 6 images, 3 appeared twice with he/him and once with they/them across lists, and 3 appeared twice with she/her and

once with they/them across lists. There were 6 names, all gender-neutral: Alex, Casey, Jaime, Jordan, Sam, Taylor (Flowers, 2015). While people who use they/them pronouns use a variety of names, this experiment uses gender-neutral names because counterbalancing the gender associations of the names within lists was not feasible. Critically, across lists they/them appears once with each image and once with each name, in order to avoid confounding interpretations about which aspects of a person's name or appearance may make it easier for someone to learn that they use they/them pronouns.

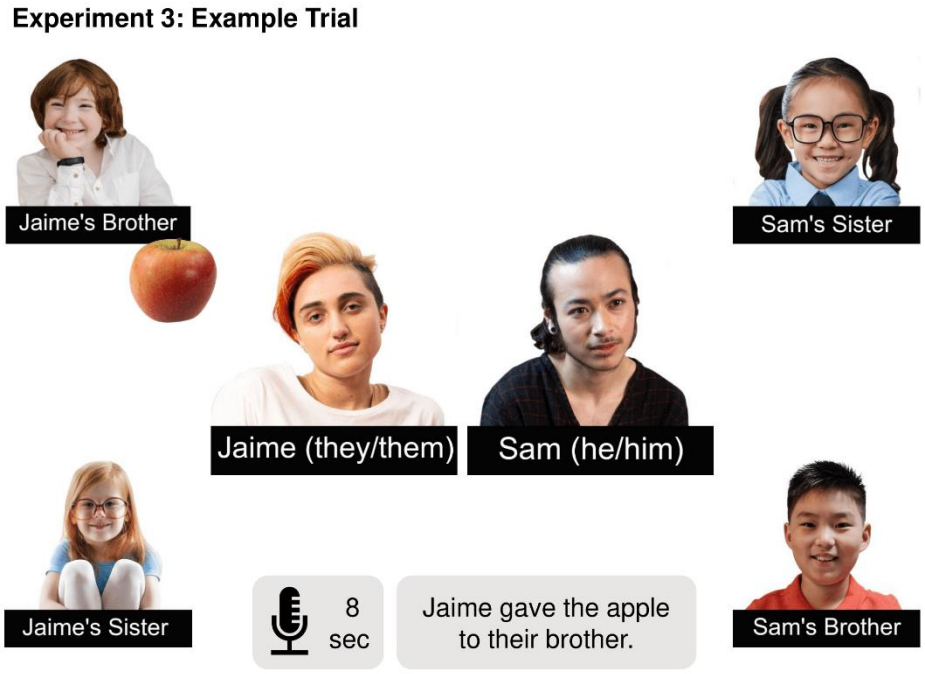


Figure 3.1: Experiment 3: Example Trial in the +Nametag condition, which preferentially elicited "Jaime gave the apple to their brother."

Pronoun Elicitation Task. Following a task established by Pozzan & Antón-Méndez (2017), each trial showed 2 characters in the center of the screen, with their siblings in the 4 corners (Figure 3.1). An animation showed an object moving from a character to one of their

siblings, and participants verbally described what happened. This was designed to elicit possessive pronouns about the target character, e.g., *Jaime gave the apple to their brother*. It also allowed participants to produce subject pronouns referring to the character, e.g., *They gave the apple to their brother*, or to avoid pronouns entirely, e.g., *Jaime gave the apple to Jaime's brother.* Trials manipulated which pronouns the two pictured characters used [Pronoun Pair]: they/them targets with he/him or she/her distractors [They|HeShe], he/him and she/her targets with they/them distractors [HeShe|They], and he/him and she/her targets with she/her or he/him distractors [HeShe|SheHe]. Trials also counterbalanced whether the object was passed to the brother or the sister and the locations of the characters. These trial frames were identical for each of the 6 character lists. Unlike Pozzan & Antón-Méndez (2017), no filler trials were included, as it was not possible to conceal that the study was targeting pronoun production.

3.2.3 Procedure

Introductions to Characters. Participants were randomly assigned to 1 of 4 between-participants conditions, manipulating what information about pronouns was given [**+Nametag vs –Nametag; +Introduction vs –Introduction**], then to 1 of 6 lists within each condition, counterbalancing the images and names of the characters. First, participants read introductions to 3 characters (1 he/him, 1 she/her, 1 they/them) (Figure 3.2). The characters were introduced by name, and in the +Introduction condition, their pronouns were explicitly stated (e.g., *This is Jaime, who uses they/them pronouns*). The images associated with each character included their name, resembling a display name in Zoom; in the +Nametag conditions, the images also included the character's pronouns in parentheses after their name. Each character had a brother and a sister, whose nametags indicated their relationship to the character. In all conditions, the facts about the character and the introductions to their siblings included 3 instances of the character's pronouns, with the character facts preceding the sibling introductions to reduce the ambiguity of singular *they* (i.e., not plural referring to the character and the sibling introduced next).

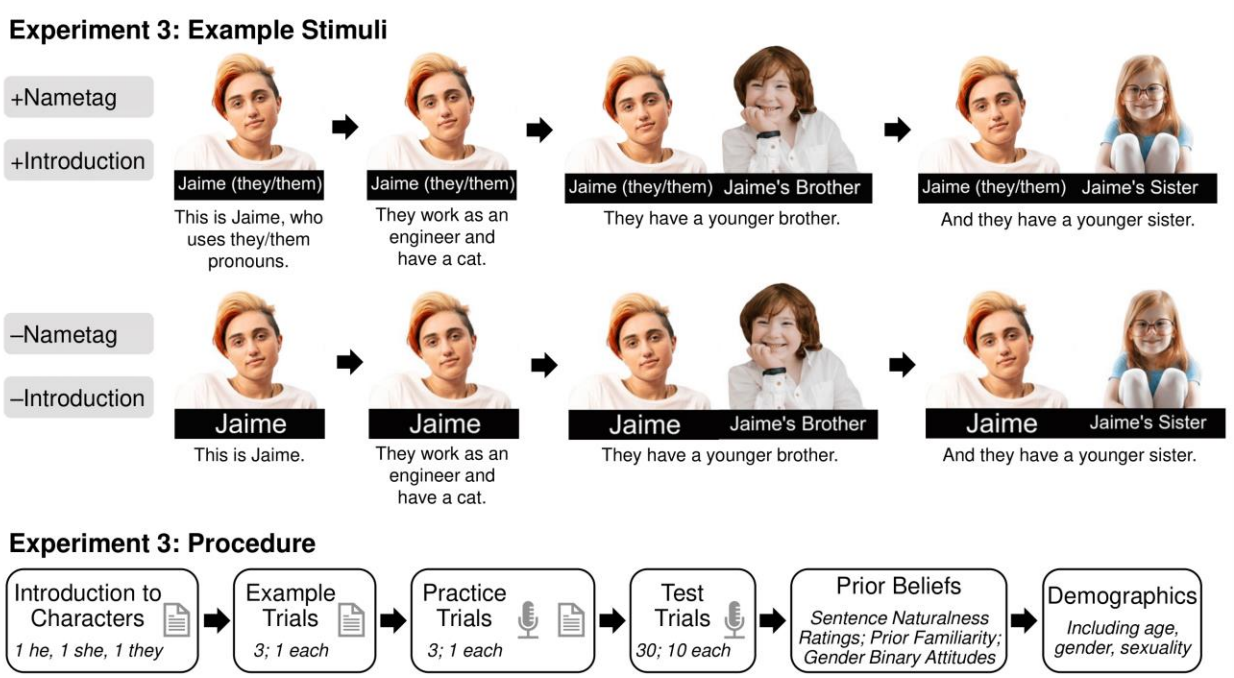


Figure 3.2: Experiment 3: Stimuli and Procedure. [A] Example stimuli for the introductions to the characters, shown for the +Nametag +Introduction and –Nametag –Introduction conditions. [B] Experiment procedure.

Speech Production. Participants saw 1 example trial for each character, which demonstrated the frame *[Name] gave the [object] to [their/his/her] [brother/sister]* and included another instance of the character's pronouns. Participants then completed 1 practice trial for each character, to learn the timing of the task. Each scene, with the object moving from beside a character to beside their brother or sister, took a total of 5 seconds. Then the microphone recorded for 8 seconds, with the images remaining on the screen. After each practice trial, participants read feedback in the frame *Did you say something like, "[Name] gave the [object] to [their/his/her] [brother/sister]?"* At this point in the experiment, participants in the condition where the characters' pronouns were not directly indicated (–Nametag –Introduction) had observed 5 examples of each characters' pronouns. Participants then completed 30 test trials, which did not include feedback. The order of all trials was randomized. Trials were divided evenly between 3 within-subjects

conditions, which varied the pronouns of both the target character and the other pictured character [Pronoun Pair: They|HeShe; HeShe|They; HeShe|SheHe].

Survey. After the speech production task, participants completed a survey measuring their prior beliefs about singular *they* and TGD identities (§B.4). First, they judged 6 sentences using singular *they*: coreferring with generic antecedents (e.g., *the ideal barista*), quantified antecedents (e.g., *each dog owner*, *every music fan*), and proper names (masculine, feminine, and gender-neutral). Items were drawn from Conrod (2019b), and participants rated them on a Likert scale with 1 being "very unnatural" and 7 being "very natural." Second, participants were asked about their prior familiarity with using they/them and pronoun-sharing practices. They could choose one or more options: use they/them for themself, close to someone who uses they/them, have met someone who uses they/them, have heard about using they/them but have not met anyone who does, and have not heard anything about using they/them. For including pronouns in introduction and in places like nametags or signatures, participants indicated frequency in the groups they were a part of (all, most, some, a few, none) and for themselves (always, usually, sometimes, rarely, never because prefer not to, never because had not heard of it). Third, participants completed Nagoshi et al.'s Transphobia Scale, which measures endorsement of gender essentialism and the gender binary, and discomfort with people who violate these expectations (2008; see also E. N. Tebbe & Moradi, 2012); this is referred to as the Gender Beliefs measure from here forward.

Finally, participants completed demographic questions (§B.1.2). They were asked about their age, gender, and sexuality, as prior work indicates that being younger and part of the LGBTQ+ community correlates with higher acceptability ratings for singular *they* (Camilliere et al., 2021; Conrod, 2019b; Hekanaho, 2020; Hernandez, 2020). The question about gender was two steps: a free-response box, then options to indicate whether their gender was the same or different than the sex indicated on their original birth certificate. This follows recommendations for identifying the broadest set of TGD people: anyone whose gender does not match their sex

assigned at birth, of whom not all call themselves transgender. This format also accounts for the fact that terms for gender vary widely, allowing participants to choose the language that best describes them, but avoiding relying on terms that many participants may not be familiar with (Ansara & Hegarty, 2014; Cameron & Stinson, 2019; NASEM, 2022; Zimman, 2017). In addition to or instead of the options for sex assigned at birth, participants could indicate whether they considered themselves cisgender, transgender, or neither. Although these factors do not relate directly to the current research questions, participants were also asked about their race, ethnicity, and education level, in order to characterize the participant sample (Buchanan et al., 2021). All demographic questions included the option to not respond. The experiment was coded and hosted using PCIbex (Zehr & Schwartz, 2018).

3.3 Predictions

Like in the first two experiments, we expect to observe lower accuracy for they/them characters compared to he/him and she/her characters, but recall that the characters in Experiment 3 differ in several ways. First, instead of masculine or feminine names, where two thirds of the characters used the expected he/him or she/her and one third used they/them, all of the names here are gender-neutral. If participants rely on lexical knowledge about the gender associations of the name to select a pronoun, responses would be split between he/him or she/her, instead of being strongly biased to one or the other. Second, the characters now include images, which provide additional information that participants may be using to make an inference about the character's gender and then to select pronouns. Third—while still brief—the introductions to the characters here contain more repetitions of the character's pronouns. While the introductions to the characters in Experiments 1 & 2 stated their pronouns directly but did not use pronouns to refer to the character (e.g., *This is Emily, who uses they/them pronouns. Emily…*), participants in all four Experiment 3 conditions see each character's pronouns used three times in the introductions, once in the example trials, and once in the practice trials. Putting

aside potential differences between spoken and written production for the moment, this means that we may see higher accuracy for they/them in the –Nametag –Introduction condition than in Experiments 1 & 2, since information about the characters' pronouns is presented multiple times.

The primary hypotheses concern whether the Nametag and Introduction conditions attenuate the lower accuracy of singular *they*. Including pronouns in introductions makes the information about pronouns salient at the beginning. If speakers use this information in production—presumably by retrieving it from episodic memory—we would expect to see a smaller penalty for they/them in the two +Introduction conditions. Alternatively, if speakers do not remember the information about the character's pronouns from the beginning of the experiment, or if this information is not used when selecting the pronoun to produce, we would see no differences between the +Introduction and –Introduction conditions.

Including pronouns on the characters' nametags keeps the information accessible throughout the experiment, and compared to the Introduction manipulation, does not require speakers to retrieve information from episodic memory. If speakers use the nametag information when selecting the pronoun to produce, we would expect to see a smaller penalty for they/them in the two +Nametag conditions. If speakers do not use the nametag information, instead relying on their lexical knowledge of the name or an inference about the character's gender based on their appearance, we would see no differences between the +Nametag and –Nametag conditions.

If introductions and nametags do reduce the relative difficulty of singular *they*, the combination may be more effective than just one, resulting in higher accuracy for the +Nametag +Introduction condition compared to the +Nametag –Introduction and –Nametag +Introduction conditions. This could result if including pronouns in introductions directs people to pay attention to the nametags, and if the nametags serve as a cue to retrieve a memory of the introduction.

3.4 Results

3.4.1 Participant Backgrounds

Participants were older than the typical college student sample (range = 19–81, Mdn = 34). Around half the participants were women, and 6 were under the nonbinary umbrella. 6 participants said that their gender was different than their sex assigned at birth and/or that they considered themselves transgender, and 38 were LGBQ+ (Table 3.1). These rates are somewhat higher than the U.S. average, but in line with previous data about the Prolific participant population (Douglas et al., 2023). Overall, just about all participants were at least somewhat familiar with singular *they* before the experiment: a third had heard about people using they/them pronouns but had not met anyone who does, a third had met but were not close to anyone who uses they/them, and a third were close to someone who uses they/them and/or used they/them themselves (Figure 3.3D). Similarly, most participants were familiar with including pronouns when introducing yourself and on nametags/display names, but didn't consider it a part of their or their social circles' norms. When describing their own habits, about half said they never do either because they prefer not to, about a quarter said they do rarely or sometimes, and about a tenth said they do usually or always (Figure 3.3B). When describing what people around them do, about a third were never around people who share pronouns, about half were rarely or sometimes around people who share pronouns, and about a fifth were usually or always around people who share pronouns (Figure 3.3C). Including pronouns on nametags was somewhat more common than including pronouns in introductions, which is unsurprising given that the former can be less marked. When rating the naturalness of singular *they* coreferring with different types of referents (Figure 3.3A), acceptance of indefinite forms was generally high ($M = 5.56$, $SD = 1.48$), and acceptance of proper names was more variable ($M = 4.58$, $SD = 1.97$) and was significantly lower ($\beta = 0.98$, $t = 4.89$, $p < .001$) (Table A3.2).

Table 3.1: Experiment 3: Participant Demographics. The trans & gender diverse and sexuality categories have variable totals, as participants could select multiple options. Participant education, English experience, and race/ethnicity are included in the appendix (Table A3.1).

Participant Demographics

Age	*162*
18–24	31
25–34	53
35–44	35
45–54	19
55–64	12
65–74	8
75+	1
Prefer not to answer / Missing data	3
Gender	*162*
Female	88
Male	63
Nonbinary	4
Genderqueer	1
Female/nonbinary	1
Woman/questioning	1
Prefer not to answer / Missing data	4
Transgender & Gender Diverse	*238*
I consider myself cisgender	63
I consider myself transgender	4
I don't consider myself cisgender or transgender	12
My gender is the same as what was written on my original birth certificate	150
My gender is different than what was written on my original birth certificate	6
Prefer not to answer / Missing data	3
Sexuality	*175*
Asexual	6
Bisexual / Pansexual	24
Gay / Lesbian	6
Heterosexual / Straight	120
Queer	9
Questioning	3
I use a different term	2
Prefer not to answer / Missing data	5
Total Participants	**162**

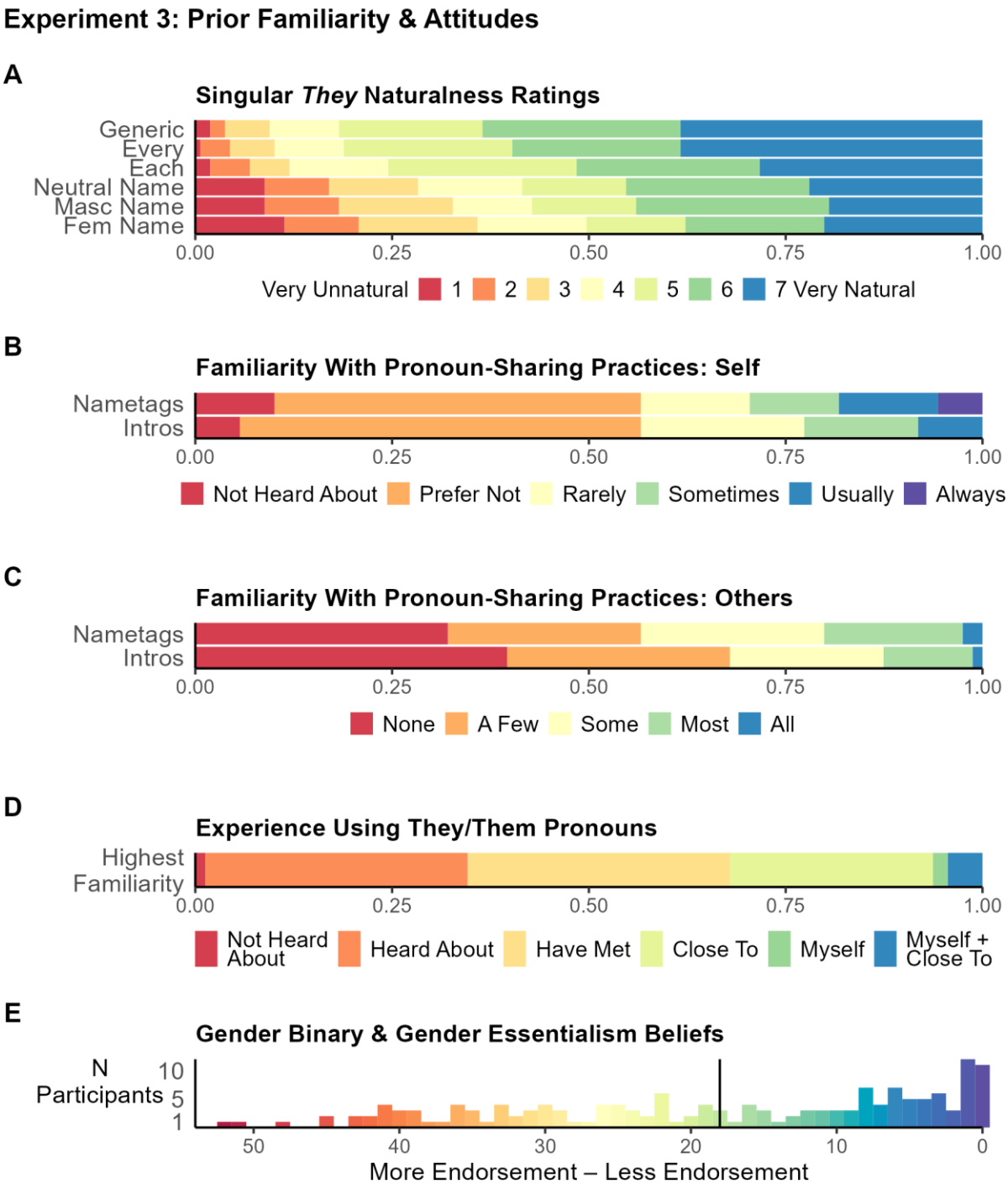


Figure 3.3: Experiment 3: Prior Familiarity and Attitudes Survey. [A] Naturalness ratings (1 = very unnatural, 7 = very natural) for singular they *coreferring with indefinites and proper names. [B] Frequency that participants include their pronouns when introducing themselves and in places like nametags. [C] Frequency that the participants' social circles include pronouns in introductions and on nametags. [D] Experience with using they/them. [E] Gender binary and essentialism beliefs, with higher scores indicating higher endorsement and thus more negative attitudes about trans and gender non-conforming people. The black line is the mean response.*

For gender beliefs (Nagoshi et al., 2008), responses on a 1–7 Likert scale were scaled to the 0–6 range, giving a total range of 0–54, with higher scores indicating higher endorsement of the gender binary and gender essentialism and thus less favorable attitudes about trans and gender-nonconforming people (Figure 3.3E). Participant totals spanned the entire scale (range = 0–52), but were skewed towards the lower end, with a mean response that was moderately favorable towards trans and gender-nonconforming people (*M* = 18.02, *SD* = 14.38; see Table A3.3 for item text and means). While this experiment did not include direct measures of political affiliation, other studies show that the Prolific population skews left, with 35% identifying as a strong Democrat, and only 20% identifying as a strong, weak, or independent Republican (Douglas et al., 2023).

3.4.2 Distribution of Pronouns Produced

Trials were automatically transcribed using *whisper* (Radford et al., 2022), then checked to include disfluencies. After excluding trials that were inaudible or did not include a response to the task (1.17% of data), 4803 trials were included in the analysis. Each trial was coded for pronoun(s) referring to the target character, which occurred in nearly all trials (98.44%). Because subject pronouns were infrequent (1 *he*, 2 *she*, 2 *they*) and did not occur in trials without a corresponding possessive pronoun, the analyses only include possessive pronouns. There were no outliers between the 6 lists varying the name-image-pronoun combinations (Figure A3.1).

Figure 3.4A shows the distribution of final pronouns produced by target character and condition. Trials with one pronoun are shown in darker colors; trials with multiple pronouns (e.g., *Jaime gave the apple to her bro—to their brother*, 1.60%) show the final pronoun in lighter colors. Participants were numerically more likely to not use pronouns for they/them characters (N = 31) than for he/him (N = 22) and she/her characters (N = 22), but the comparison between the mean rates of no-pronoun responses for they/them and he/him + she/her was not statistically significant, $t(2747.18) = -1.42$, $p = .16$. Rather, participants who only used names (e.g., *Jaime gave the apple to Jaime's brother*) tended to do so for all 3 characters.

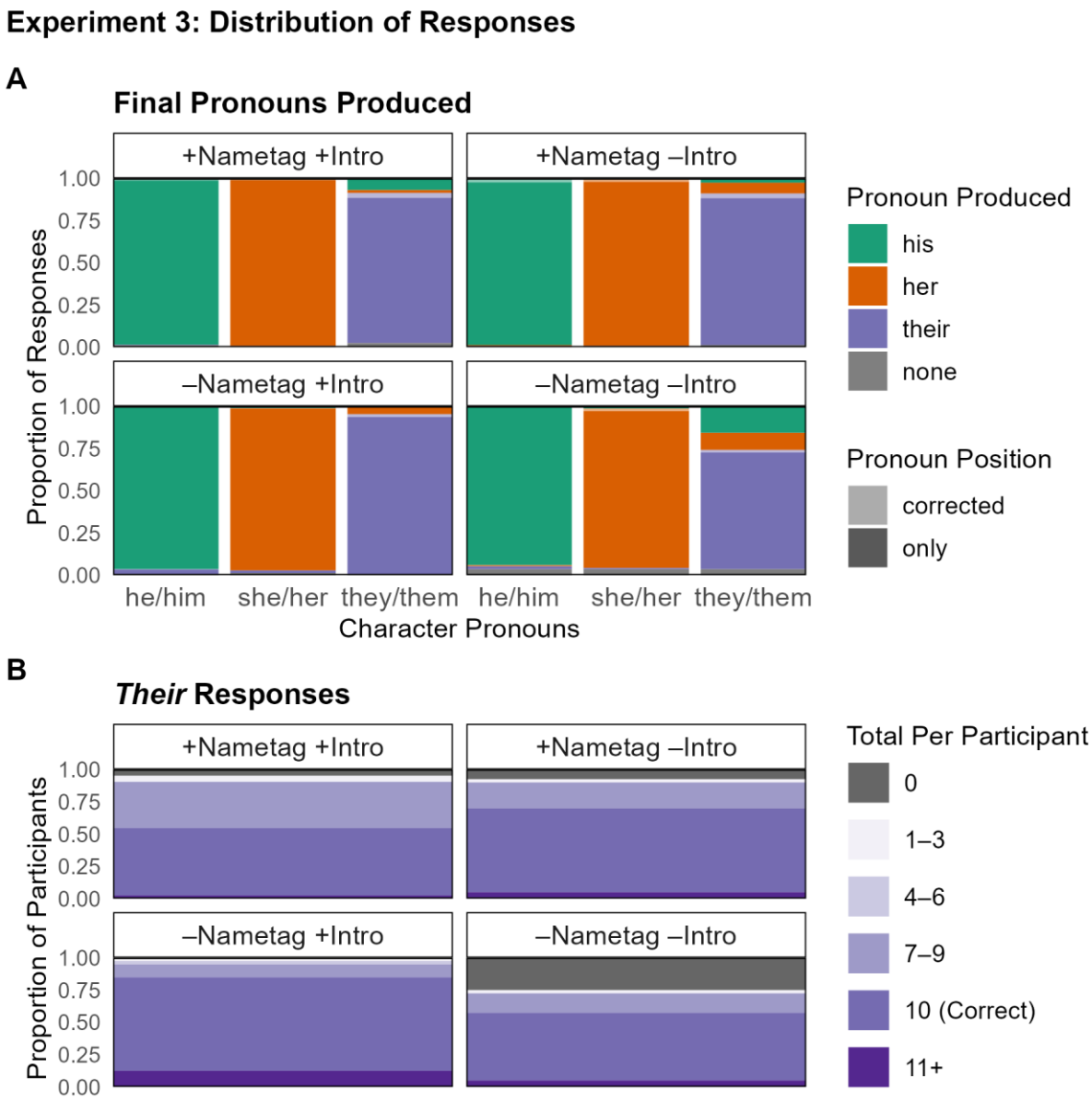


Figure 3.4: Experiment 3: Distribution of Responses. [A] Final pronoun produced. Trials where participants used multiple pronouns to refer to the character (e.g., "Jaime gave the apple to her bro—to their brother") are grouped based on the final pronoun and shown in lighter colors. [B] Number of their *responses per participant.*

In trials where participants produced multiple pronouns, self-corrections from *his* to *their* (N = 14) and *her* to *their* (N = 23) were more common than self-corrections from *their* to *his* (N = 5) or *her* (N = 8), or between *his* and *her* (N = 12, N = 15). Unexpectedly, *their* responses for each participant were typically at floor or near ceiling (Figure 3.4B), with the Nametag and Introduction conditions affecting whether a participant produced *their* at all, more than affecting accuracy within participants who produced *their* in some trials.

3.4.3 Pronoun Accuracy

The primary analysis measured accuracy of pronouns referring to the target character (Figure 3.5). Trials where participants produced different pronouns (1.60%) were coded based on the final pronoun (e.g., *Jaime gave the apple to her bro—to their brother* would be coded based on the accuracy of *their*); trials with no pronouns (1.56%) were excluded from this analysis. Pronoun was manipulated between both the character described (target) and the other character pictured on the screen (distractor). The first Pronoun Pair contrast compared trials with they/them target characters to trials with he/him and she/her target characters (They|HeShe vs HeShe|They + HeShe|They). Within he/him and she/her character trials, the second contrast compared trials with they/them distractors to trials with he/him and she/her distractors (HeShe|They vs HeShe|They). There were no significant effects of the second Pronoun Pair contrast, so for simplicity's sake, the first Pronoun Pair contrast is referred to as the effect of Pronoun from here on. The fixed effects of Nametag and Introduction (between-participants) were both mean-center effects coded. The maximal model justified by the experimental design (Baayen et al., 2008; Barr et al., 2013) included by-participant slopes for Pronoun and by-item intercepts. Item was defined as the combinations of names, images, and pronouns that varied across the 6 lists of characters; because these combinations did not fully vary across pronouns, by-item random slopes were not included. The final model (Table 3.2) included all interactions between fixed effects, plus by-item and by-participant intercepts (Bates et al., 2015; R Core Team, 2023; Voeten, 2023).

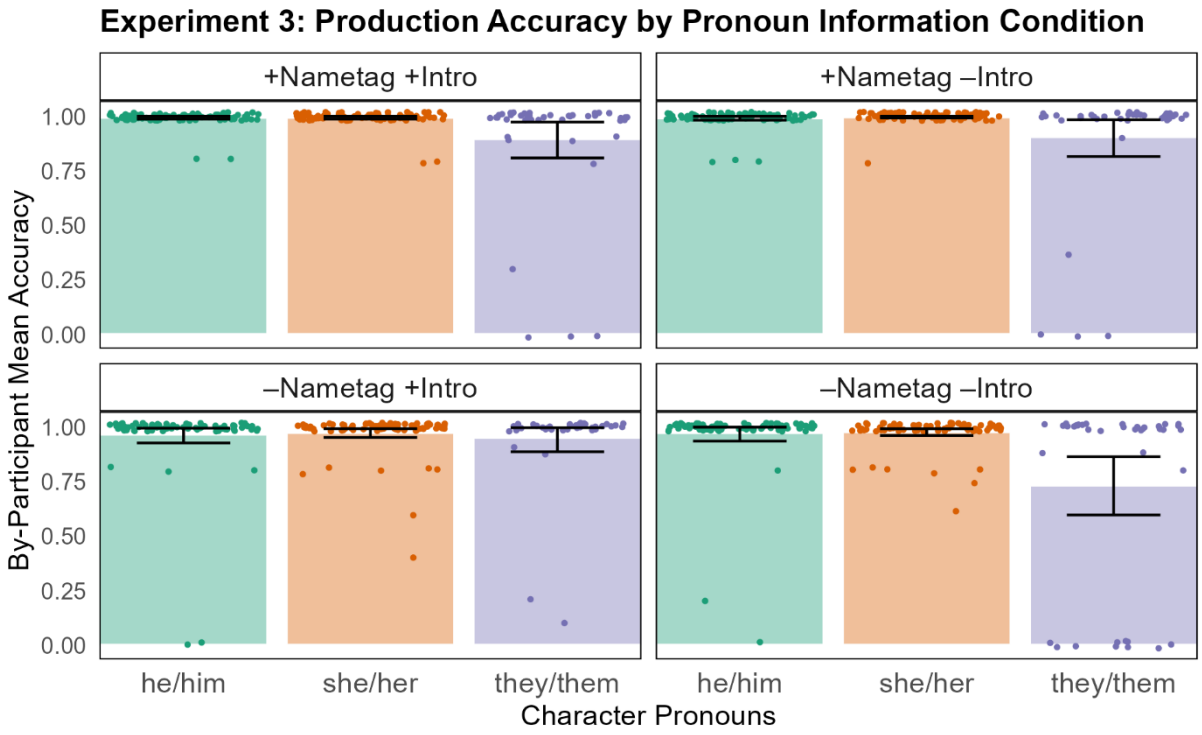


Figure 3.5: Experiment 3: Production Accuracy, split by Nametag and Introduction conditions. Trials using no pronouns are excluded. By-participant means are shown as points; error bars indicate 95% Cis calculated over the by-participant means.

Across all conditions, participants were more likely to produce the correct pronoun than not (β = 13.16, z = 12.26, p < .001). Participants were more accurate for he/him and she/her characters (M = 0.98 across Nametag and Introduction conditions) than for they/them characters (M = 0.88) (β = 5.05, z = 5.09, p < .001). Within he/him and she/her trials, there was no significant difference between trials where the other pictured character used he/him or she/her and trials where the other character used they/them (β = -0.76, z = -1.61, p = .11). The main effects of Nametag and Introduction were not significant (β = 1.71, z = 1.54, p = .12; β = -0.80, z = -0.72, p = .47). The three-way interaction between Pronoun, Nametag, and Introduction was significant (β = 6.24, z = 3.92, p < .001). This was qualified by significant interactions between Pronoun and Introduction (β = -2.58, z = -3.36, p < .001), and Pronoun and Nametag (β = 2.32, z = 2.91, p < .01). Post-hoc comparisons showed that Introduction attenuated the difference in accuracy

between they/them and he/him + she/her in the –Nametag conditions (β = -5.70, z = -6.93, p < .001), but not in the +Nametag conditions (β = 0.54, z = 0.41, p = .68). Figure 3.6 shows the means for each condition: accuracy for he/him and she/her characters was near ceiling for all conditions, and accuracy for they/them characters was highest in the –Nametag +Introduction condition (M = 0.95, SD = 0.21), slightly lower in the +Nametag +Introduction (M = 0.91, SD = 0.28) and +Nametag –Introduction (M = 0.91, SD = 0.29) conditions, and lowest in the –Nametag –Introduction condition (M = 0.73, SD = 0.44).

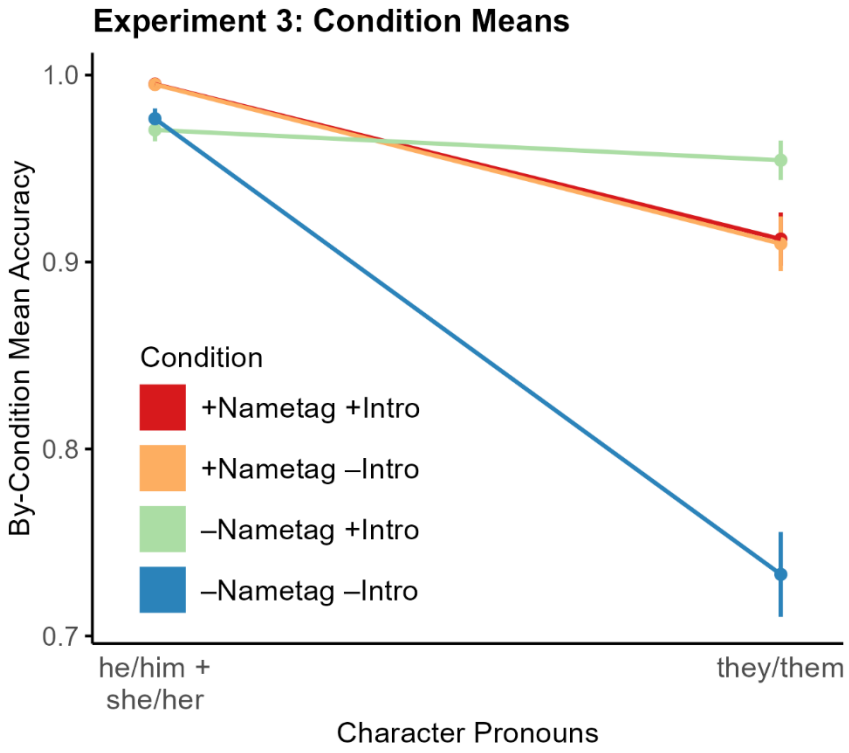


Figure 3.6: Experiment 3: Condition Means. Means and 95% CIs of accuracy for he/him + she/her characters and they/them characters, split by Nametag and Introduction conditions.

Table 3.2: Experiment 3: Production Accuracy. Model results for the effects of Pronoun Pair, Nametag, and Introduction on Pronoun Accuracy (=1), with trials that did not include a pronoun referring to the target character excluded, and trials that contained different pronouns coded based on the final one.

Experiment 3: Production Accuracy

	Production Accuracy			
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	13.162	1.073	12.263	**<0.001**
Pronoun Pair: They|HeShe (-.66) **vs HeShe|They** (+.33) **+ HeShe|SheHe** (+.33)	5.047	0.991	5.092	**<0.001**
Pronoun Pair: HeShe|They (-.5) vs HeShe|SheHe (+.5)	-0.763	0.475	-1.606	0.108
Nametag (-.5; +.5)	1.706	1.108	1.540	0.124
Introduction (-.5; +.5)	-0.797	1.110	-0.718	0.473
Pronoun (They|HeShe vs HeShe|They + HeShe|SheHe) *** Nametag**	2.324	0.798	2.912	**0.004**
Pronoun (HeShe|They vs HeShe|SheHe) * Nametag	-1.045	0.951	-1.099	0.272
Pronoun (They|HeShe vs HeShe|They + HeShe|SheHe) *** Introduction**	-2.575	0.766	-3.360	**0.001**
Pronoun (HeShe|They vs HeShe|SheHe) *Introduction	-0.294	0.951	-0.309	0.757
Nametag * Introduction	1.296	2.259	0.574	0.566
Pronoun (They|HeShe vs HeShe|They + HeShe|SheHe) *** Nametag * Introduction**	6.240	1.591	3.923	**<0.001**
Pronoun (HeShe|They vs HeShe|SheHe) * Nametag * Introduction	0.324	1.902	0.170	0.865
Random Effects				
$T_{00\ Participant}$	71.775			
$T_{00\ Character}$	2.841			
$N_{Participant}$	161			
$N_{Character}$	18			
Observations	4729			

3.4.4 Exploratory Analyses

To estimate internal reliability, I used the Bayesian mixed-effects model approach described in Staub (2021). The trials were split in half, so that each half of the data included 5 he/him, 5 she/her, and 5 they/them characters for each participant. Pronoun was coded as 2 separate variables: the first comparing they/them (-.66) to he/him (+.33) and she/her (+.33) in even trials, with odd trials coded as 0, and the second comparing they/them to he/him + she/her in odd trials, with even trials coded as 0. The *brms* package in R (Bürkner, 2017) fit a model with the odd and even trial Pronoun variables as fixed effects predicting accuracy and as by-participant random slopes. The model kept the default priors and was fit using 4 chains, each with 4000 iterations, of which 2000 were warm-up. The random slope estimates represent the relative accuracy of they/them compared to he/him + she/her for each participant, and these estimates were strongly correlated between halves of the data, $r = 0.97$ [0.90, 1.00]. This matches the distribution of results for they/them characters (Figure 3.4B), where participants tended to produce singular *they* in all or nearly all trials, or in none.

After confirming that the task showed high internal reliability, I conducted exploratory analyses with participant covariates that have previously been shown to correlate with acceptability ratings for singular *they.* For the sentence naturalness ratings (1–7 with 7 as "very natural"), the mean ratings for the generic, *each*, and *every* sentences [**Indefinite Ratings**] and for the 3 proper name sentences [**Name Ratings**] were calculated for each participant. For experience using they/them pronouns [**Familiarity**], participants were split into 3 similar-sized groups: had not heard about it before the study or had heard about it but hadn't met anyone who does (N = 55); had met someone who uses they/them but weren't close to them (N = 53); used they/them pronouns themself and/or were close to someone who uses they/them (N = 51). For familiarity with pronoun-sharing practices [**Sharing**], "none" and "never, because I had not heard about this" responses were coded as 0; "never, because I prefer not to" responses were coded as 1; "rarely" and "a few" responses were coded as 2; "sometimes" and "some" responses were

coded as 3; “usually” and “most” responses were coded as 4; and “always” and “all” were coded as 5. These 4 questions were summed to create 1 composite score, with 0 indicating the lowest familiarity and 20 indicating the highest (M = 7.24, SD = 4.85). Responses for the gender binary and gender essentialism beliefs measure [**Gender Beliefs**] were rescaled from 1–7 to 0–6 and summed, with higher responses indicating stronger endorsement. Sexuality was coded as 1 for participants who said they were asexual, bisexual/pansexual, gay/lesbian, and/or queer (N = 38) and 0 otherwise (N = 121). Because only 6 participants said they were transgender and/or that their gender is different than their sex assigned at birth, analyzing this as a separate factor is difficult. Instead, **LGBTQ+** identity is treated as one variable, noting that the 6 TGD participants were also LGBQ+.

The strongest correlation between participant covariates (Figure 3.7) was familiarity with pronoun-sharing practices and familiarity with using they/them pronouns (r = 0.62, $p < .001$), which is unsurprising given that people who use they/them nearly always have to explicitly state their pronouns in order to not be misgendered. The brief naturalness ratings questionnaire largely replicated prior results correlating gender beliefs, familiarity, LGBTQ+ identity, and age with judgments about singular *they* (Camilliere et al., 2021; Conrod, 2019b; Hekanaho, 2020; Hernandez, 2020; Minkin & Brown, 2021; Nichols et al., 2019; Parker et al., 2019). The second-strongest correlation was between naturalness ratings and gender beliefs, with participants who more strongly endorsed the gender binary and gender essentialism rating singular *they* coreferring with proper names as less natural (r = -0.52, $p < .001$) (see in particular Hernandez, 2020). LGBTQ+ participants (r = 0.30, $p < .001$) and participants more familiar with using they/them (r = 0.33, $p < .001$) and with pronoun-sharing practices (r = 0.38, $p < .001$) rated *they* coreferring with proper names as more natural. Older participants rated it as less natural (r = -0.22, $p < .01$). However, ratings for indefinite singular *they* (generic, *each*, *every*) were not significantly correlated with other participant covariates—or with ratings for *they* coreferring with proper names (r = 0.07, $p = .40$).

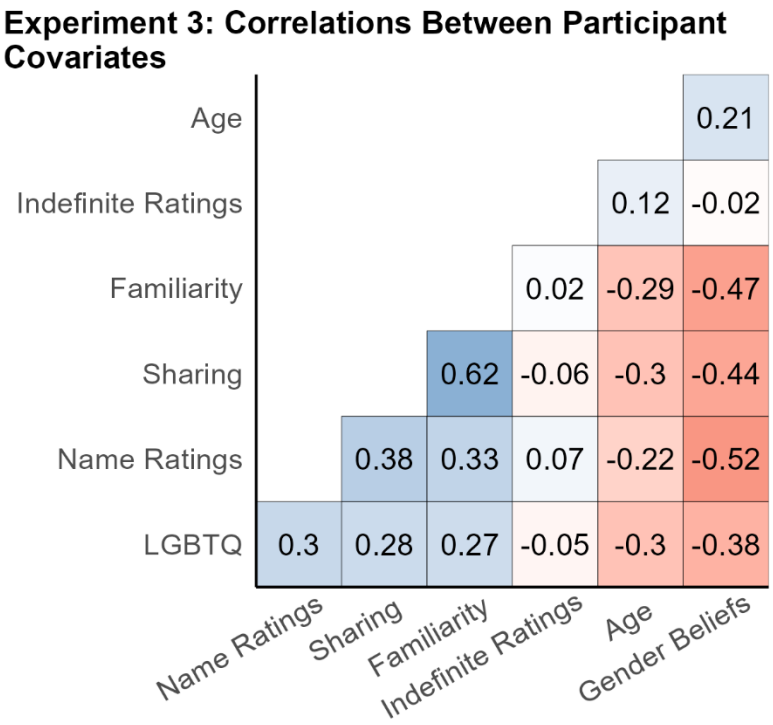


Figure 3.7: Experiment 3: Correlations Between Participant Covariates. Age, LGBTQ+ identity, familiarity with using they/them pronouns and pronoun-sharing practices, naturalness ratings for singular they *coreferring with indefinite referents and with proper names.*

I then tested if adding these participant covariates to the hypothesis-testing model significantly contributed to fit. Age, Familiarity, Gender Beliefs, Name Ratings, and Sharing were mean-centered; and LGBTQ+ was mean-center effects coded. The distributions of the rescaled variables are shown in Figure A3.2. The *buildmer* package in R (Bates et al., 2015; R Core Team, 2023; Voeten, 2023) was used to identify the most complex converging model, allowing all interactions between fixed effects. It then performed backwards stepwise elimination to remove participant covariate terms that did not significantly contribute to model fit, while retaining all of the fixed and random effects from the hypothesis-testing model. The final model included Gender Beliefs, Familiarity, and a subset of their two- and three-way interactions (Table A3.4). No effects of Familiarity were significant after Bonferroni correction for multiple comparisons.

Participants who more strongly endorsed the gender binary and gender essentialism were less accurate overall (β = -10.07, z = -3.44, p < .001) and showed a larger relative difference in accuracy between they/them and he/him + she/her (β = 6.50, z = 3.53, p < .001) (Figure 3.8). The interaction between Pronoun, Introduction, and Gender Beliefs was marginally significant after correction for multiple comparisons (β = 13.32, z = 3.11, p < .01), such that Gender Beliefs had a larger effect on the relative accuracy of they/them in the –Introduction conditions than in the +Introduction conditions.

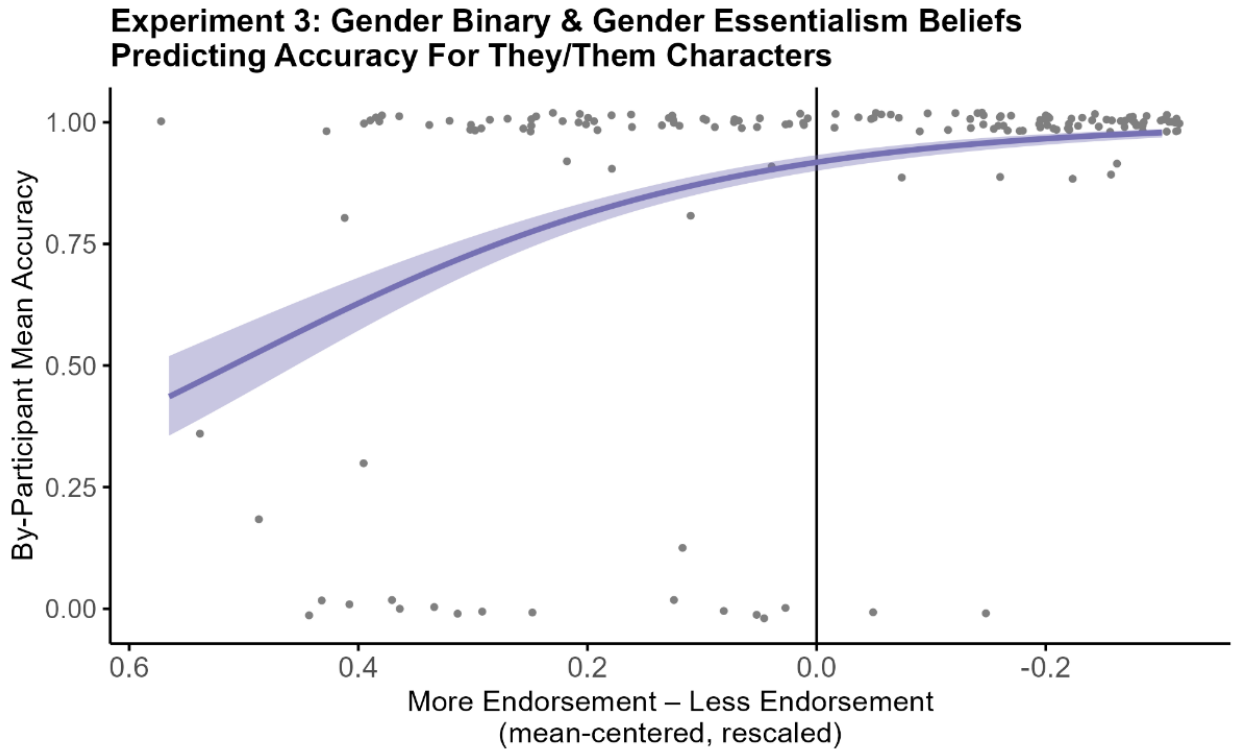


Figure 3.8: Experiment 3: Accuracy by Gender Beliefs. By-participant mean accuracy for they/them characters, predicted by endorsement of the gender binary and gender essentialism. Points are by-participant means; the line is a GLM fit over the raw data.

3.5 Discussion

In Experiment 3, participants learned about three characters, each of whom was associated with pronouns (1 he/him, 1 she/her, 1 they/them), a gender-neutral name, an image, and two sibling images. In all conditions, participants saw a total of five examples of each character's pronouns in use before beginning the test trials. Additional information about the characters' pronouns varied by two factors: the Introduction conditions manipulated whether the introductions to the characters explicitly stated *who uses __ pronouns*, and the Nametag conditions manipulated whether the images of the character included pronouns alongside their name. In each trial, participants saw two characters in the center, with their four siblings in the corners. An object moved from a character to a sibling, prompting spoken descriptions in the frame *Jaime gave the apple to their brother* (Pozzan & Antón-Méndez, 2017). This structure preferentially elicited—but did not require—participants to produce a possessive pronoun.

Baseline accuracy for they/them characters was high compared to the first two experiments, with participants in the –Nametag –Introduction condition correctly producing singular *they* in about three quarters of trials. Both the nametag and introduction manipulations facilitated singular *they*, with accuracy in conditions with one or both rising to over 90%. Accuracy for they/them characters was, unexpectedly, highest in the –Nametag +Introduction condition and slightly lower in the +Nametag +Introduction and +Nametag –Introduction conditions. Generally, the Nametag and Introduction conditions tended to affect whether or not participants used singular *they* at all, with the majority of participants producing singular *they* in all or nearly all trials, or in no trials.

The original goal of this experiment was to investigate how introductions and nametags may reduce the number of errors speakers make—potentially mirroring the real-life situation where well-intentioned people do get they/them pronouns correct, but only when paying attention to it, and will frequently default back to he/him or she/her when the demands of the conversation direct their attention elsewhere. The speech production task proved relatively easy for

participants, since they did not have to remember names for the characters and their siblings, the trial pacing erred on the side of not requiring them to rush, and the objects were typically easy to name. While it would not have been possible to conceal the fact that the experiment is about pronouns, it is likely that producing pronouns was the most difficult aspect of the task, and that participants were able to focus their attention on it. From this perspective, the task proved too easy for participants.

However, the all-or-nothing distribution of responses means that the task showed high internal reliability, warranting individual differences analyses. Participants were recruited from Prolific and had a wide range of experience using they/them pronouns, naturalness ratings for various forms of singular *they* (Conrod, 2019b), experiences with pronoun-sharing practices, and beliefs about the gender binary and gender essentialism (Nagoshi et al., 2008). This experiment replicates the expected relationships between acceptability judgments for *they* coreferring with proper names. Age and endorsement of the gender binary negatively correlated with naturalness ratings, and LGBTQ+ identity, experience with they/them, and familiarity with pronoun-sharing positively correlated with naturalness ratings (Camilliere et al., 2021; Conrod, 2019b; Hekanaho, 2020; Hernandez, 2020; Minkin & Brown, 2021; Parker et al., 2019). However, when testing if adding participant covariates to the hypothesis-testing model improved fit, acceptability ratings did *not* predict production accuracy. Instead, gender beliefs was the strongest predictor of accuracy for they/them characters, with participants who more strongly endorsed the gender binary and gender essentialism and expressed more discomfort with gender non-conforming people being less likely to use singular *they.*

Collecting speech production data online came with both advantages and drawbacks. Compared to in-lab experiments, recruiting participants on Prolific resulted in a sample more diverse in age, language experience, and sociopolitical beliefs (Douglas et al., 2023), as well as the ability to collect enough data to be appropriately powered. The fact that participants never interacted directly with an experimenter means that social desirability pressures may have been

different. While it was clear that the experiment wanted them to use they/them pronouns, speakers may feel differently about refusing or failing to do so when an addressee is present. How the social context—either another participant completing the same task, or a researcher who participants may infer is LGBTQ+ and is invested in the outcome of the experiment—is an area for future research.

CHAPTER 4

Experiment 4: Online Processing

4.1 Motivation

One of the most common objections to singular *they* is that both generic and specific forms are too ambiguous (Hekanaho, 2020). However, it is unclear whether this opinion arises from actual difficulties in coreference resolution, or if it is more a product of language and gender attitudes. According to the processing fluency account (Alter & Oppenheimer, 2009), processing difficulty and language attitudes are connected both directly and indirectly. First, processing fluency can be a cue to language attitudes: if a listener attributes their difficulty understanding a speaker to that speaker being unable or unwilling to communicate in a way that the listener finds clear, experiencing less processing fluency will cause them to evaluate the speaker more negatively. Second, harder processing tends to elicit more negative affect, which may bias listeners' language attitudes (Dragojevic, 2020). One line of experiments has connected the processing fluency account specifically to perceptions of nonnative-accented speech. Participants listened to audio recordings of fictional stories, and while their task was ostensibly to remember enough of the story to complete a fill-in-the-blanks memory task, the dependent measures were how much processing fluency they experienced (e.g., rating as clear, easy to understand), how positively they felt about the speaker (1–100 scale), and status (e.g., intelligence, competence) and solidarity (e.g., friendliness, niceness) judgments about the speaker. The experiments manipulated various ways of making the audio easier or harder to understand, independent of the speaker's accent. Adding background white noise to the audio decreased listeners' fluency ratings when the speaker used a Punjabi accent, more so than when the same speaker used a Standard American English accent. The lower processing fluency then resulted in more negative affect and lower status attributions to the speaker. When a Mandarin-accented speaker was accompanied

with subtitles or participants had read a transcript of the story first, participants reported higher processing fluency, which resulted in more positive feelings about and higher status attributions to the speaker. In both sets of experiments, the effects of making the listening conditions easier or harder on status attributions were mediated by processing fluency and sequentially by fluency and affect (Dragojevic, 2020; Dragojevic et al., 2017; Dragojevic & Giles, 2016).

The processing fluency account would predict that dislike of singular *they* is caused, at least in part, by lower-level processing difficulty. Multiple factors could cause listeners to experience lower processing fluency for singular *they* compared to other pronouns: a larger set of possible antecedents may make it more ambiguous, it may elicit a number or gender mismatch agreement violation, it may be newly learned, and it is overall less frequent even for speakers familiar with it. The processing fluency account also predicts that making singular *they* easier to understand would reduce people's negative reactions to it, and therefore the people who use it.

However, the actual amount of processing difficulty for singular *they*—particularly for definite specific gender-specified forms—is unclear. Only a few studies to date have investigated online comprehension of *they* coreferring with proper names. These are described in more detail in §0.4.4, but to review briefly, people are slower to identify the referent for *they* compared to *he* and *she*, as measured through a maze task while reading (Shenkar et al., 2023) and a mouse tracking task while listening (Arnold et al., 2023). In two ERP experiments, a P600 effect was observed for *they* coreferring with proper names (gender-specified), but not with specific gender-unspecified referents (e.g., *the participant*) (Chen et al., 2023; Prasad & Morris, 2020). Since the P600 indexes detecting a syntactic error or having difficulty comprehending a sentence's syntactic structure (Hagoort et al., 1993; Kaan et al., 2000; Osterhout et al., 1994; Osterhout & Holcomb, 1992), Prasad & Morris interpret their results to indicate that *they* coreferring with proper names still causes a gender agreement error, even though participants in their experiment all had significant experience with using they/them pronouns and considered it grammatical in offline acceptability judgments.

However, results like Prasad & Morris (2020) do not necessarily require that *they* for specific gender-specified antecedents is still ungrammatical for these participants. Even in LGBTQ+ communities, *they* coreferring with a name is still relatively infrequent overall and would not be expected in many contexts. Stimuli in sentence processing experiments are typically unrelated, with each sentence using a different name and referring to a new character. When singular *they* corefers with a new referent in each trial, it is unclear whether *they* is consistently perceived as syntactically anomalous, or if it is originally unexpected, but could be processed smoothly once anticipated to corefer with a particular referent. Experiment 4 tests processing in the context of repeated reference, where listeners can come to expect singular *they* to corefer with certain characters. This is potentially somewhat easier, and it more closely resembles the real-world contexts in which we hear pronouns referring to people.

Additionally, the majority of processing studies, particularly for generic indefinite *they*, have used self-paced reading and eyetracking while reading measures. Experiment 4 is one of the first to use the visual world paradigm, which measures eye movements while participants listen to sentences describing a visual scene. Gaze at pictured characters provides a measure of online processing as the sentence unfolds, since listeners automatically look at what they think is being talked about (Allopenna et al., 1998; Sedivy et al., 1999; Spivey et al., 2002; Tanenhaus et al., 1995, 2000). The visual world paradigm has advantages compared to other tasks, as it provides detailed time-course information about *which* alternative interpretations are being considered, in addition to *when* processing difficulties occur.

Experiment 4 investigates the degree of processing difficulty of singular *they* compared to *he* and *she*, if the processing of singular *they* follows the same patterns as *he* and *she*, and if processing measures correspond with offline judgments. The design is based on a prior line of work investigating ambiguous pronoun resolution. In Arnold et al. (2000, 2007), participants looked at illustrated scenes of cartoon characters and listened to stories about them:

[21] *1 Donald is bringing some mail to Mickey/Minnie, while a violent storm is beginning.*
2 He's/She's carrying an
3 umbrella and it looks like they're both going to need it.

Part 1 introduced 2 named characters (*Donald, Mickey/Minnie*), using a verb (*bringing*) that allows for a subsequent pronoun to refer to either of the characters individually (Garnham, 2001; Gordon et al., 1993; Sanford & Garrod, 1981). While *he* or *she* (part 2 in [21]) is more likely to refer to the character mentioned first in the prior sentence (*Donald*) (Arnold et al., 2000, 2007; Gernsbacher, 1989; Kaiser & Trueswell, 2011), it can also refer to the character mentioned second (*Mickey/Minnie*). In other words, the character mentioned first is more accessible (Ariel, 2006). This structure makes it possible for the referent of *he* or *she*—called the target character in visual world experiments—to remain ambiguous until the next phrase (part 3 in [21]) can be compared to the illustration. In this example, either Donald or Mickey/Minnie is carrying an umbrella (Figure 4.1A). This allows for enough time to observe processing of the pronoun (*is carrying an*), but without creating a discourse context too different from actual language use.

The stories in Arnold et al. (2000) manipulated 2 factors: the ambiguity of the pronoun (target and competitor characters using the same vs different pronouns) and the accessibility of the referent (target mentioned first vs second). The results showed that listeners rapidly use both gender and accessibility cues to identify which character the pronoun referred to (Figure 4.1B). If gender was unambiguous (top right in Figure 4.1A), the pronoun referred to the character mentioned first (bottom left), or both (top left), participants looked at the target character starting at approximately 200ms after the pronoun. This is about as quickly as effects in the visual world paradigm can be observed (Hallett, 1986; Tanenhaus et al., 1995). When neither gender nor accessibility cues disambiguated the referent (bottom right), participants looked at the target and competitor characters almost equally. For the purposes of the present experiment, these results provide a validated stimuli design and a baseline for how we expect *he* and *she* to be processed.

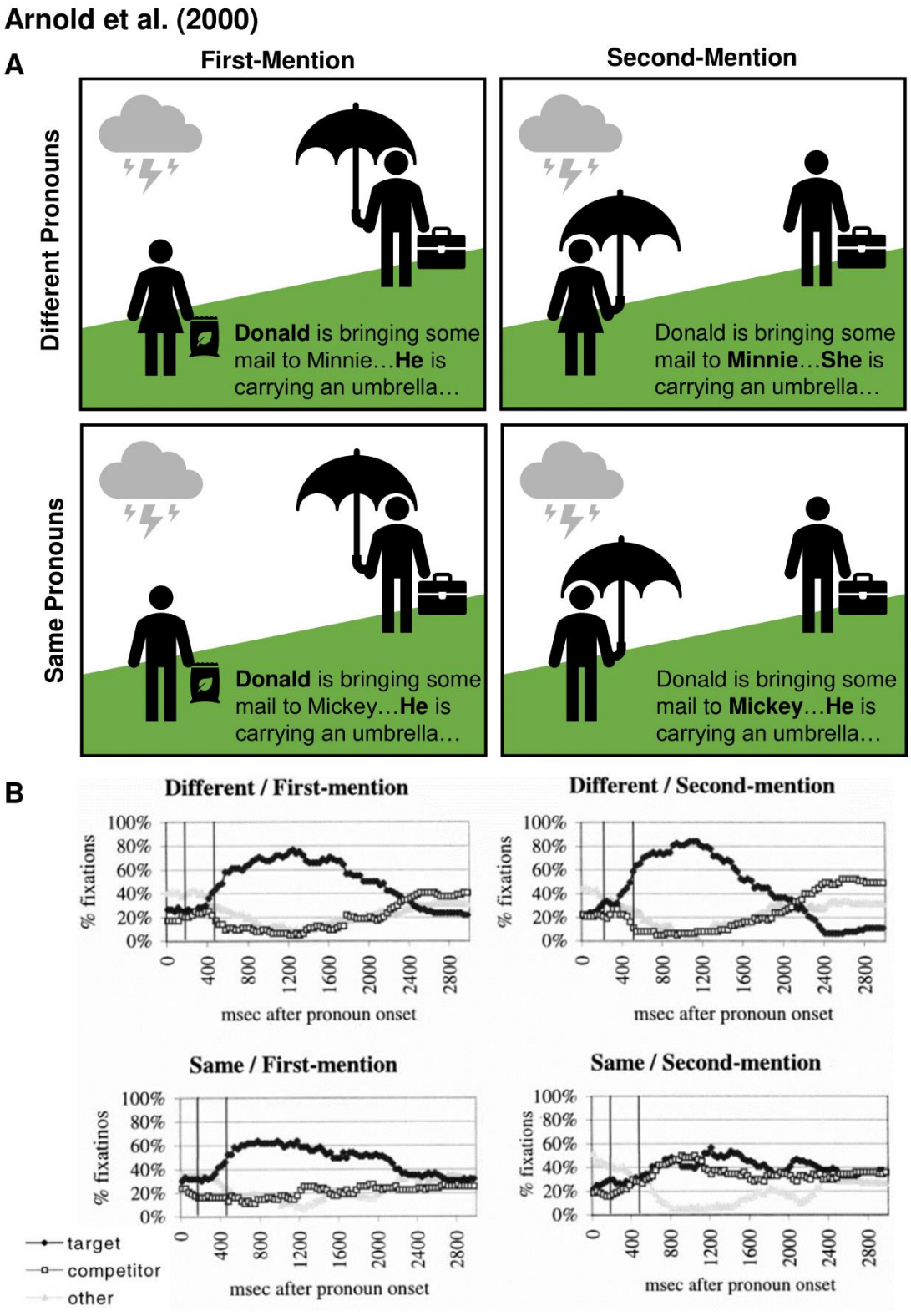


Figure 4.1: Arnold et al. (2000). [A] Recreation of design, showing the pronoun ambiguity and order of mention conditions. The original materials were illustrated using the Disney characters. [B] Results, with 0 indicating pronoun onset and horizontal lines indicating the verb (carrying).

A later set of studies used a similar design to examine how listeners process pronouns acoustically ambiguous between *he* and *she* (Brown-Schmidt & Toscano, 2017; Falandays et al., 2020). This experiment used similar stories as Arnold et al. (2000, 2007), but different images. Instead of 2 characters being drawn to match the scene, characters were pictured in colored shapes. The target character was disambiguated by describing their location, e.g., *he's standing on a blue square* instead of *he's carrying an umbrella*. This allows for a larger range of stimuli to be created, and the prior results demonstrate that listeners can process *he* and *she* smoothly in these types of stories. Critically, while the descriptions may seem odd and somewhat discontinuous, the discourse structure matches how speakers introduce new referents and when they tend to use pronouns instead of names.

The current experiment uses similar manipulations as Arnold et al. (2000, 2007) and the same task as Brown-Schmidt & Toscano (2017). One potential issue with this design is that *they* can be ambiguous between a singular and plural interpretation, even if participants learn that *they* is always singular in the context of the experiment. An alternative is to use stimuli that rule out a plural interpretation of *they*. Reflexive pronouns (*himself, herself, themself*) can syntactically constrain a singular interpretation (e.g., Runner et al., 2006; Sturt, 2003), but introduce a potential confound, since speakers vary in whether they prefer *themself* or *themselves* for singular referents (Ahn & Conrod, 2022). Another option is to use stimuli that semantically rule out a plural interpretation. Returning to some of the examples in the first chapter (§0.2.3), *they're worrying* [6] can be singular or plural, since people can worry together, but *their free leg* [7] can only be singular, since a body part only belongs to one person. However, it is difficult to create stimuli that rule out a plural interpretation, while still including a long enough period where the pronoun is ambiguous between two possible referents. Results like these would be difficult to interpret because it would be unclear whether processing costs are due to singular *they* itself, or because the structure of the story does not match when speakers use pronouns instead of names or other referring expressions. Moreover, because fully ruling out a plural interpretation is difficult, the

majority of instances of singular *they* in actual language use *do* contain some degree of ambiguity between singular and plural interpretations. Results from stimuli that reflect a very narrow set of contexts in which people hear singular *they*, where no plural interpretation is at all possible, would be less relevant.

4.2 Methods

The design and analysis plan were preregistered on the Open Science Framework (10.17605/OSF.IO/R3FY9). Sources and attributions for the images are located in the appendix (§B.5.2); the edited images and audio stimuli are available upon request. The de-identified data and analysis code are available at this dissertation's Github repository.

4.2.1 Participants

30 participants completed the study for partial course credit or for pay; their demographic information is shown in Table 4.1. Participants were required to be fluent English speakers (but not necessarily native or monolingual) and to have normal or corrected-to-normal vision and hearing, and most were Vanderbilt undergraduate students. An additional 2 participants completed the experiment, but were excluded due to too few trials having usable eyetracking data. The experiment lasted approximately 45 minutes.

4.2.2 Materials

Characters. Participants learned about 6 characters, each associated with a name and an image: 2 who used he/him, 2 who used she/her, and 2 who used they/them (Figure 4.2A). The 6 character names and 6 character images were the same as in Experiment 3 (Drucker, 2019) (§B.5.2). Recall that all names were gender neutral since counterbalancing gender associations of the names within lists was not feasible. Participants were randomly assigned to 1 of 6 lists, in order to counterbalance the images and names associated with characters who use they/them.

Across lists, 3 images appeared twice with he/him and once with they/them, and 3 images appeared twice with she/her and once with they/them; each name appeared twice with each pronoun. Critically, across lists they/them appeared once with each image and once with each name, in order to avoid confounding interpretations about what aspects of a person's name or appearance may make it easier for someone to learn that they use they/them pronouns.

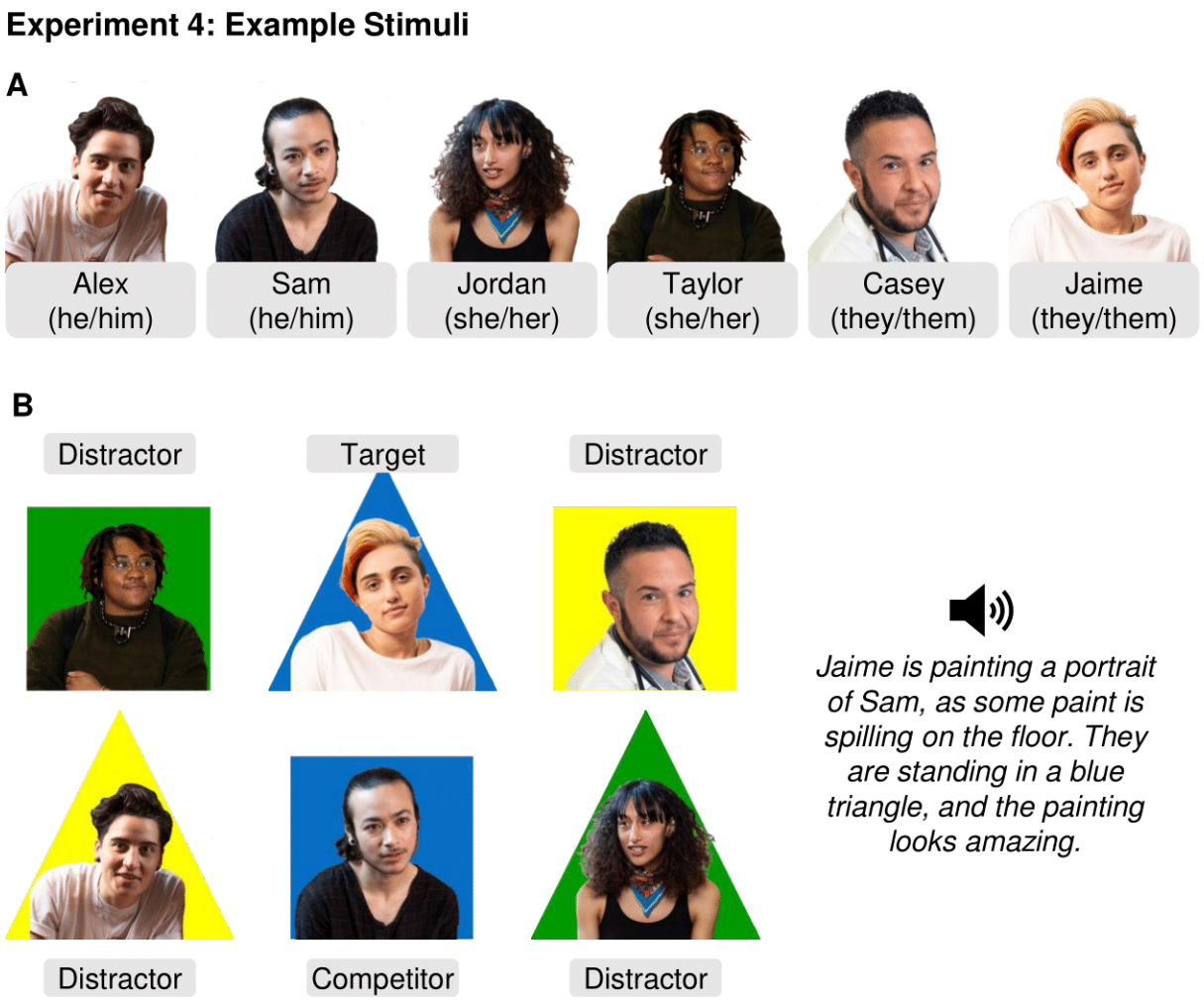


Figure 4.2: Experiment 4: Stimuli. [A] Example set of characters. [B] Example trial screen and story, with grey boxes indicating information not shown to participants.

Stories. The stories and visual scenes were based on Arnold et al. (2000, 2007) and Brown-Schmidt & Toscano (2017). During each trial, the 6 characters were arranged in a 3x2 grid, each shown inside a colored shape (red, yellow, green, blue; triangle, square) (Figure 4.2B). Participants listened to stories in the frame:

[22] *1 Jaime is painting a portrait of Sam, as some paint is spilling on the floor.*
2 He is/she is/they are
3 standing in a blue triangle
4 and the painting looks amazing

Each story began with a sentence that named two characters, with an additional phrase to allow time for participants to identify them (part 1 in [22]). The two named characters—the target and competitor—always used different pronouns (e.g., Jaime: they/them, Sam: he/him). This created 3 **Pronoun Pair** conditions: they/them targets with he/him or she/her competitors [**They|HeShe**], he/him or she/her targets with they/them competitors [**HeShe|They**], and he/him or she/her targets with he/him or she/her competitors [**HeShe|SheHe**]. Next, a pronoun (*he, she,* or *they*) referred to one of the named characters (part 2 in [22]). This created 2 **Order of Mention** conditions, where the pronoun refers to the character mentioned **first** in the preceding sentence or to the character mentioned **second**. Figure 4.2B shows an example of the first-mention condition, where the pronoun (*they*) refers to the first named character (Jaime). The second-mention story matching this scene would have *he is standing in a blue triangle*, where *he* refers to Sam.

At this point in the story, participants could identify which of the named characters is the target if they knew the characters' pronouns and were using that information in their language comprehension (e.g., Jaime uses they/them and Sam uses he/him, meaning that *they* refers to Jaime). The stories then described the location of the target character (part 3 in [22]). Because the target and competitor characters were always pictured with the same color, the target was not fully disambiguated until the shape word, an average of 1364ms after the pronoun onset. After

the shape word, listeners could identify the target character without taking the pronoun into consideration. The story concluded with a final phrase, which did not include another pronoun referring to the character(s) (part 4 in [22]). After listening to each story, participants were asked to decide whether it matched the scene (e.g., if Jaime was standing in a blue square).

There were a total of 60 story frames (1 + 4 in [22]) (§B.6). Within lists, each story appeared once in the first-mention condition and once in the second-mention condition. Across lists, each story appeared twice with each pronoun for counterbalancing, but with the same pair of names to make the stimuli recording feasible. There were a total of 24 pronoun + color + shape combinations (2 + 3 in [22]). These clips were recorded as full sentences (not spliced together), and each trial randomly selected 1 of 3 versions, in order to avoid participants learning additional cues about a particular recording. The audio was recorded by the first author, a white native English speaker from the northeast U.S. with a feminine voice.

4.2.3 Procedure

Character Learning. To learn about the characters, participants first saw each character's image, accompanied by their name (e.g., *This is Jaime*) and a fact about them (e.g., *They like to play the piano, They work as an engineer*). Each character was shown twice, so that participants saw two examples of the characters' pronouns. However, pronouns were never directly stated (e.g., *This is Jaime, who uses they/them pronouns*), and the use of singular *they* was not explained to participants. Participants were then tested on the names and images of the characters. They were shown all 6 images and asked to click on the named character. If the answer was correct, the image and name of the character was displayed, along with another example of their pronouns (e.g., *Correct, they're Jaime*). If the answer was incorrect, the image of and information about the incorrectly chosen character was shown (e.g., *Incorrect, he's Sam*), followed by the image of and information about the correct character (e.g., *They're Jaime*). To continue, participants were required to get all 6 names correct in the same block. When listening

to the stories, participants should then have been able to identify the images of the 2 named characters, and had seen at least 3 examples of each characters' pronouns.

Eyetracking. During each trial, the images were displayed for 1 second, then the audio began playing. After the story finished, the images remained on the screen, and the text *Did the story match the picture?* was displayed at the bottom. Participants clicked *YES* or *NO* at the corner of the screen to advance to the next trial. Eye movements were recorded with an Eyelink 1000 desktop-mounted eyetracker recording monocularly at 1000 Hz, with drift correction after every fifth trial. The trial order was randomly generated for each participant, the locations of the 6 images were randomly generated for each trial, and the colors and shapes were counterbalanced.

Participants completed 6 practice trials in order to explain the task and that they should judge whether the story matched the scene based on the colored shape sentence, since the action described at the beginning (e.g., painting a portrait) was not pictured. These practice trials used a name instead of a pronoun (e.g., *Jaime is standing in a blue triangle*). 4 trials matched the scene, and 2 trials mismatched by referring to a color not pictured. After each practice trial, participants saw feedback on if their match judgment was correct.

Participants then completed 96 critical and 18 filler trials, mixed in a randomized order. These varied according to 2 within-subjects factors: Pronoun Pair [They|HeShe; HeShe|They; HeShe|SheHe] and Order of Mention [target mentioned first; second]. The target and competitor characters were evenly distributed, yielding a total of 32 critical trials for each pronoun. Filler trials were included to ensure that participants treated *no* as an option in the match judgment question, even if they considered singular *they* acceptable and knew which characters used they/them. 10 of the filler trials were unambiguously the wrong description, referring to a color that was not pictured on the screen (e.g., for Figure 4.2B, *they are standing in a red square*). The other 8 filler trials used one of the pronouns of the non-named characters, making the story incorrect for the target character, as well as the competitor character (e.g., for Figure 4.2B, *she is standing in a blue triangle*). The he/him and she/her characters were each called *they* twice, and the they/them

characters were each called *he* once and *she* once. No filler trials used *he* instead of *she* or *she* instead of *he*. Note that throughout the experiment, *they* was always singular, never plural. After completing the 120 trials, participants were tested on the names of the characters, following the same procedure as before, but without feedback.

Survey. Finally, participants completed the same singular *they* naturalness ratings, familiarity with using they/them pronouns, gender binary and gender essentialism beliefs (§B.4), and demographics questions (§B.1.2) as in Experiment 3. All demographic questions included the option to not respond. Figure 4.3 shows an overview of the full procedure.

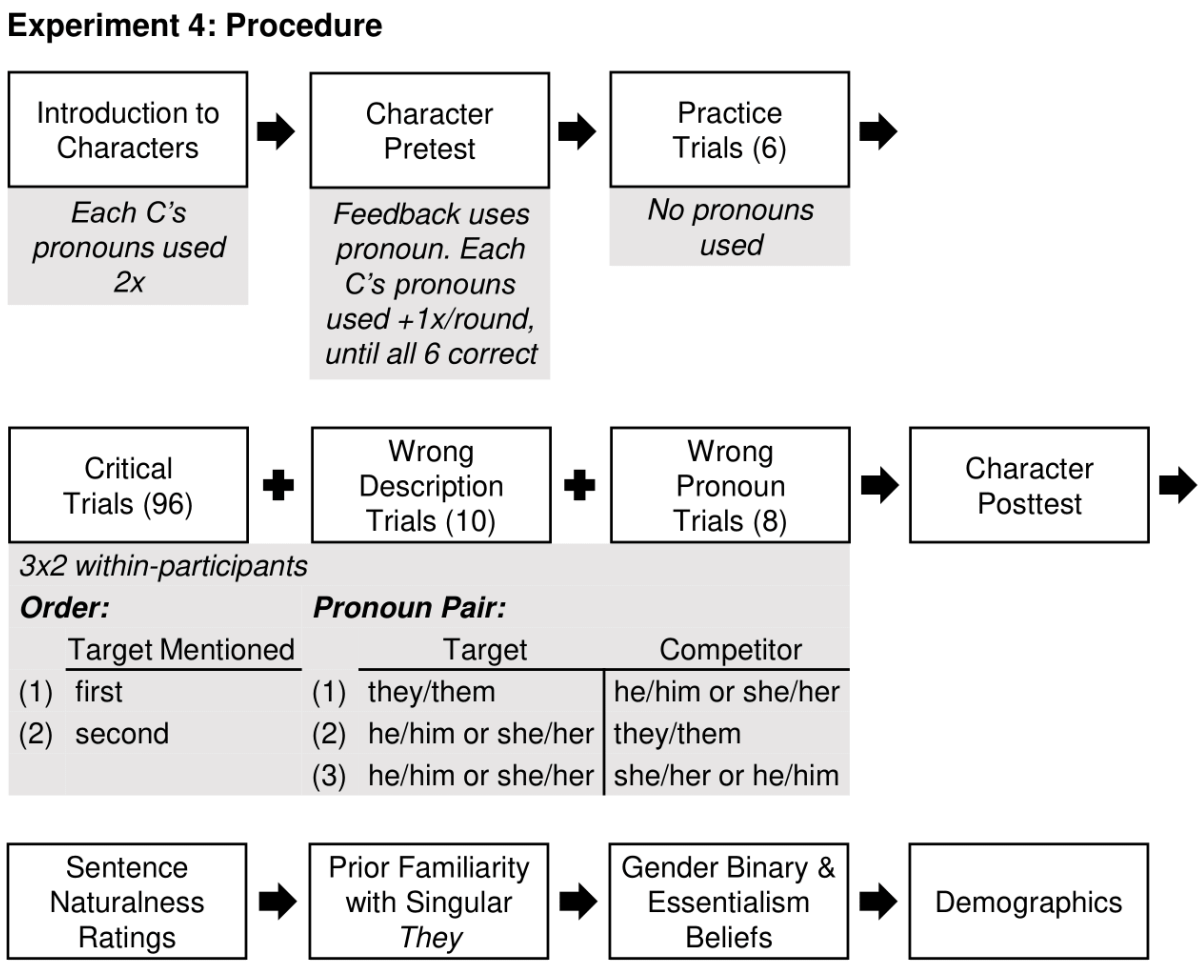


Figure 4.3: Experiment 4: Procedure.

4.3 Predictions

The first question concerns whether listeners can accurately comprehend *they* as singular, then combine this with knowledge about the character's pronouns to identify who is being described in the story. If so, participants will preferentially look at the target character after the pronoun and before the disambiguating shape word. While it is theoretically possible that singular *they* would show no processing costs compared to *he* or *she*, prior results indicate this is currently unlikely (Arnold et al., 2023; Chen et al., 2023; Prasad & Morris, 2020; Sanford & Filik, 2007; Shenkar et al., 2023). Instead, listeners may identify the referent for singular *they* before the disambiguation, but more slowly than they do for *he* and *she*. This result would resemble those observed in young children (Arnold et al., 2007; Song & Fisher, 2005) and in adult second language learners (Cunnings et al., 2017; Grüter et al., 2012; Speyer & Schleef, 2019), who can use gender and order of mention cues from pronouns to identify the referent, but do so more slowly than fluent adults. Alternatively, there are two ways of observing results where listeners do not preferentially look at the target before the disambiguating shape word, which the current experiment cannot distinguish between. One possibility is that listeners attempt to use singular *they* to identify the target character, but do not succeed because of ambiguity. Another possibility is that listeners recognize the potential ambiguity in *they* and strategically choose to wait until hearing more information before deciding on an interpretation.

A secondary question concerns the competitor character, who is named at the beginning of the story but whose pronouns are never used. Stories using *he* and *she* can have a competitor who uses she/her or he/him (never the same as the target character), or a competitor who uses they/them. A difference between these two conditions could be predicted in either direction: If trials where the competitor character uses they/them are slower than trials where the competitor character uses he/him or she/her, this could indicate that some aspect of the they/them characters—the pronoun activated alongside the character or the character themself—is causing greater competition (making it a stronger possibility) than the he/him and she/her characters. If,

on the other hand, trials where the competitor character uses they/them are faster, this could indicate that listeners are treating the more ambiguous character as less likely to be referred to, either in general or with a pronoun.

With regards to order of mention, we expect to replicate prior results for *he* and *she*, where participants are more likely to look at target characters who were named first than target characters who were named second, because although the pronoun can refer to either, it is more likely to refer to the person mentioned first (Arnold et al., 2000, 2007; Brown-Schmidt & Toscano, 2017). If we also observe an order of mention effect for singular *they*—either the same as or present but reduced compared to *he* and *she*—it would suggest that singular *they* is being integrated into listeners' standard discourse processing mechanisms.

4.4 Results

4.4.1 Participant Backgrounds

To contextualize the findings, I first discuss the results of the survey. Most participants were in the typical undergraduate age range (M = 20.13, SD = 2.61) and described themselves as native English speakers (N = 27). 19 were women, 11 were men, and none identified as transgender and/or a gender different than their sex assigned at birth (Table 4.1). Overall, all participants were at least somewhat familiar with singular *they* before the experiment: 5 had heard about people using they/them pronouns but not met anyone who does, 21 had met but were not close to anyone who uses they/them, and 9 were close to someone who uses they/them, but no participants used they/them themselves (Figure 4.4B). When rating the naturalness of singular *they* coreferring with different types of referents (Figure 4.4A), acceptance of indefinite forms was generally high (M = 5.17, SD = 1.76). Surprisingly, ratings for proper names (M = 4.98, SD = 1.66) were not significantly lower than ratings for indefinites (β = 0.19, t = 0.36, p = .73) (Table A4.2). For the gender beliefs measure (Nagoshi et al., 2008), responses were again scaled

to 0–6 and summed, so that a score of 0 indicated the lowest endorsement of the gender binary and gender essentialism, and a score of 54 indicated the highest (Figure 4.4C). Participant totals spanned the entire range but were strongly skewed towards the lower end, with the mean response favorable towards trans and gender-nonconforming people (range = 1–53, *M* = 13.93, *SD* = 12.56) (see Table A4.1 for item text and means).

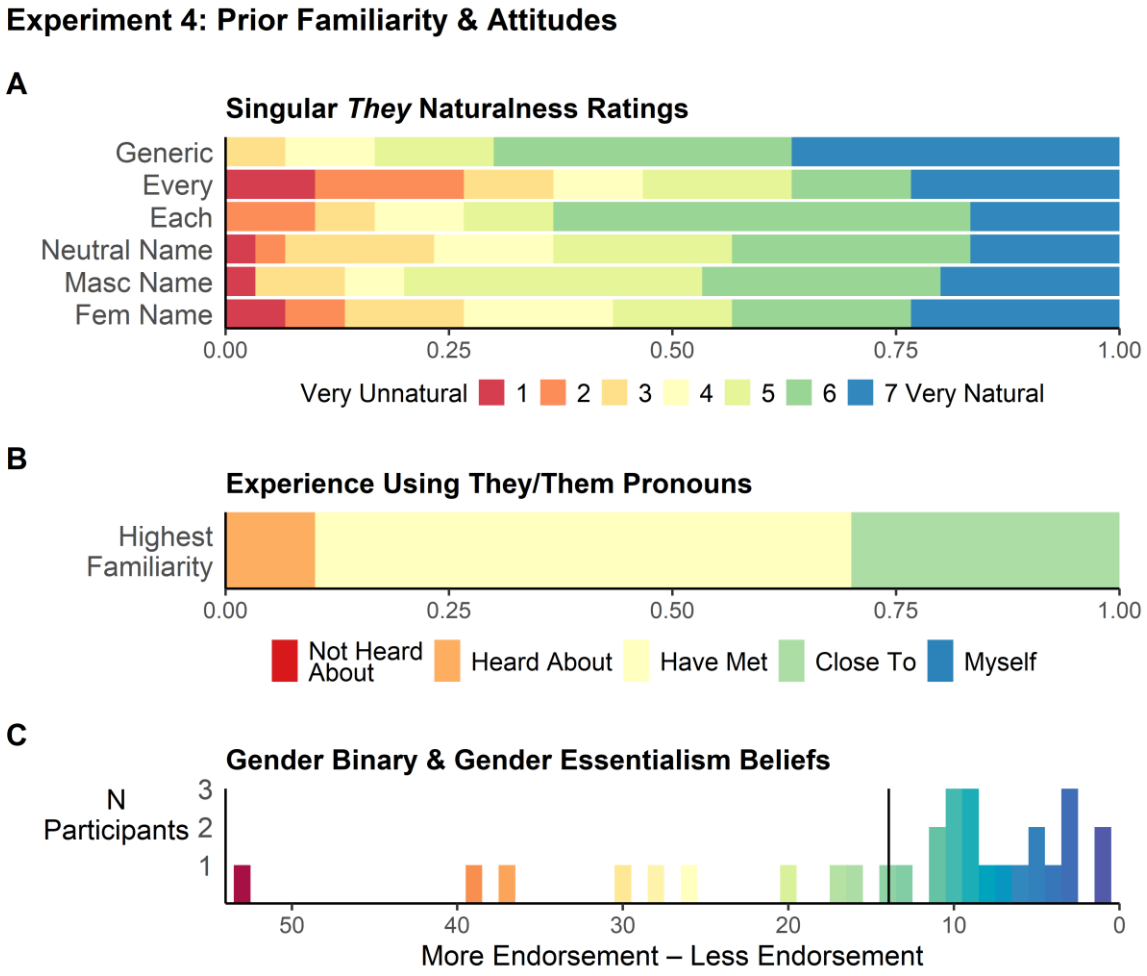


Figure 4.4: Experiment 4: Prior Familiarity and Attitudes Survey. [A] Naturalness ratings on a 7-point Likert scale (1 = very unnatural, 7 = very natural) for singular they coreferring with indefinite referents and with proper names. [B] Experience with using they/them pronouns. [C] Gender binary and essentialism beliefs, with higher scores indicating higher endorsement and thus more negative attitudes towards transgender and gender non-conforming people (Nagoshi et al., 2008). The mean response is indicated by the black line.

Table 4.1: Experiment 4: Participant Demographics. Categories with higher totals allowed participants to select as many options as applied. All questions included the option to not respond.

Participant Demographics

Age	*30*
18	10
19	6
20	4
21	6
25	1
26	2
27	1
Gender	*30*
Female	19
Male	11
Transgender & Gender Diverse	*47*
I consider myself cisgender	17
I consider myself transgender	0
I don't consider myself cisgender or transgender	0
My gender is the same as what was written on my original birth certificate	30
My gender is different than what was written on my original birth certificate	0
Sexuality	*34*
Asexual	1
Bisexual/Pansexual	3
Gay/Lesbian	1
Heterosexual/Straight	27
Queer	0
Questioning	2
I use a different term	0
English Experience	*30*
Native (learned from birth)	27
Fully competent in speaking, listening, reading, and writing, but not native	3
Race/Ethnicity	*34*
American Indian or Alaska Native	0
Asian	10
Black, African American, or African	1
Hispanic, Latino, or Spanish	6
Middle Eastern or North African	1
Native Hawaiian or Pacific Islander	0
White	15
I use a different term	0
Prefer not to answer	1
Total Participants	**30**

4.4.2 Offline Measures

Character Learning. Participants were generally able to learn the name-image pairs within 2–3 test rounds (*M* = 2.33, *SD* = 1.09). Across all pretest rounds, accuracies for they/them characters (*M* = 0.81, *SD* = 0.40) and she/her characters (*M* = 0.83, *SD* = 0.38) were slightly lower than accuracy for he/him characters (*M* = 0.90, *SD* = 0.30). Participants remembered the names of the characters throughout the study, with most (N = 27) getting all 6 correct in the post-test, and no participants excluded for getting 4 or fewer correct.

Match Judgments. When asked if the description they heard matched the scene (Figure 4.5, left), participants correctly judged the majority of test trials to be matching. The match rates for singular *they* trials (*M* = 0.91) were not significantly lower than the match rates for *he* and *she* trials ($M_{HeShe|They}$ = 0.91, $M_{HeShe|SheHe}$ = 0.94) (Table A4.3). For the wrong description trials, which referred to a color that was not pictured, participants were correctly at floor for all pronouns. For the wrong pronoun trials, which used the pronoun that neither of the two named characters used, participants varied. However, they were not less likely to indicate a mismatch when they/them characters were referred to with *he* or *she* (*M* = 0.44) than when he/him or she/her characters were referred to with *they* (*M* = 0.43) (Table A4.4). Reaction times were calculated from the display of the match question until the participant's click, with responses outside of 3SD of the mean of each trial type excluded as outliers (Figure 4.5, right). Similar to the accuracy data, reaction times were shortest for wrong description trials (*M* = 2917ms, *SD* = 2360ms), somewhat longer for test trials (*M* = 3544ms, *SD* = 3369ms), and longest and most variable for wrong pronoun trials (*M* = 4995ms, *SD* = 3996ms). An inverse Gaussian distribution with an identity link was fit to test whether Pronoun Pair affected reaction time in the test trials (Table A4.5). This accounts for the non-normal distribution of reaction time data, but in contrast to applying a non-linear transformation (e.g., log), it maintains the theoretical assumption that experimental manipulations affect the total amount of time to make a decision (Lo & Andrews, 2015).

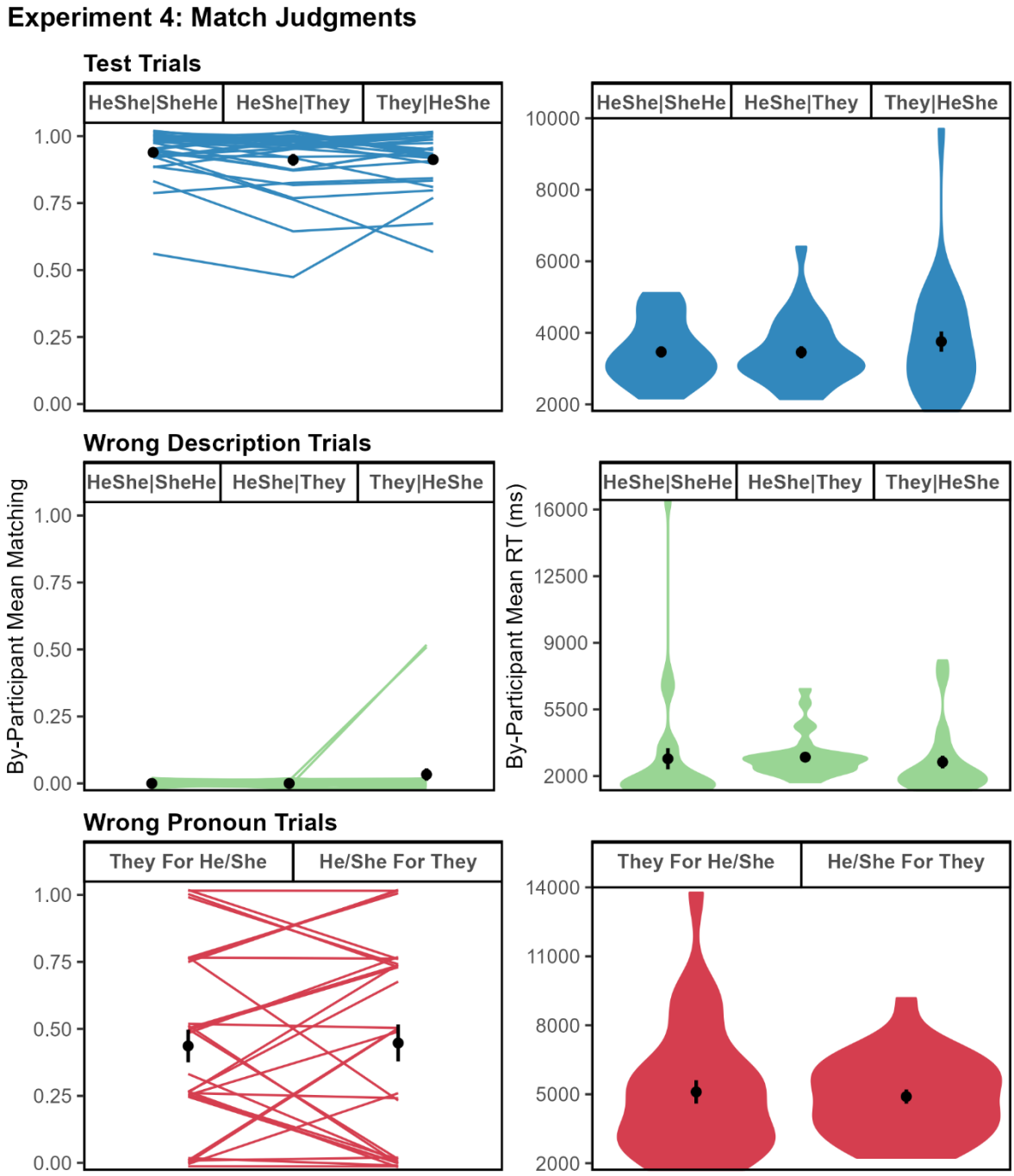


Figure 4.5: Experiment 4: Match Judgments. By-participant mean proportions of stories judged to match the picture (left) and reaction times (right). Lines indicate by-participant means between Pronoun Pair conditions; violins indicate the distribution of by-participant means; point ranges indicate condition means and 95% CIs calculated over the by-participant means. The correct answers are match (=1) for test trials and mismatch (=0) for wrong pronoun and wrong description trials. For the wrong description trials, he/him or she/her for a they/them character corresponds to the HeShe|SheHe condition; they/them for a he/him or she/her character corresponds to the They|HeShe condition; and there were no wrong pronoun trials for the HeShe|They condition.

The maximal model that converged included by-participant and by-item intercepts and slopes for Pronoun Pair (Bates et al., 2015; R Core Team, 2023; Voeten, 2023). Participants were slower to make match judgments for stories using singular *they* than stories using *he* and *she* (β = -506.67, t = -12.40, p < .001). The pronoun of the competitor character did not affect reaction times (β = 32.96, t = 0.75, p = .45).

4.4.3 Online Processing

Figure 4.6 shows fixations to the target, competitor, distractor, and no characters, starting 500ms before the onset of the pronoun and continuing for 2500ms (e.g., *…spilling on the floor. They are standing in a blue triangle, and the painting looks amazing*). The He|She and She|He trials (first row) generally resemble prior results (Arnold et al., 2000, 2007; Brown-Schmidt & Toscano, 2017), with participants starting to look at the target more than the competitor rapidly after the onset of the pronoun, and more so when the target character was mentioned first. Unexpectedly, the order of mention effect—where listeners look more at the character named first in the story than to the character named second—is only clear in He|She trials, not in She|He trials. The He|They and She|They trials (second row) show the expected order effect, but participants are less likely to be looking at the target than in the He|She and She|He trials. The They|He and They|She trials (third row) still show participants looking at the target more than the competitor before the onset of the shape word, but less than in the other two conditions, and no order effect is apparent. Examining fixations during the beginning of the story (Figure A4.1), participants looked at the target and competitor after each was named, and the time course did not differ by the character's pronouns. This confirms that participants knew the names of the characters and had identified the two possible referents before the start of the critical time window.

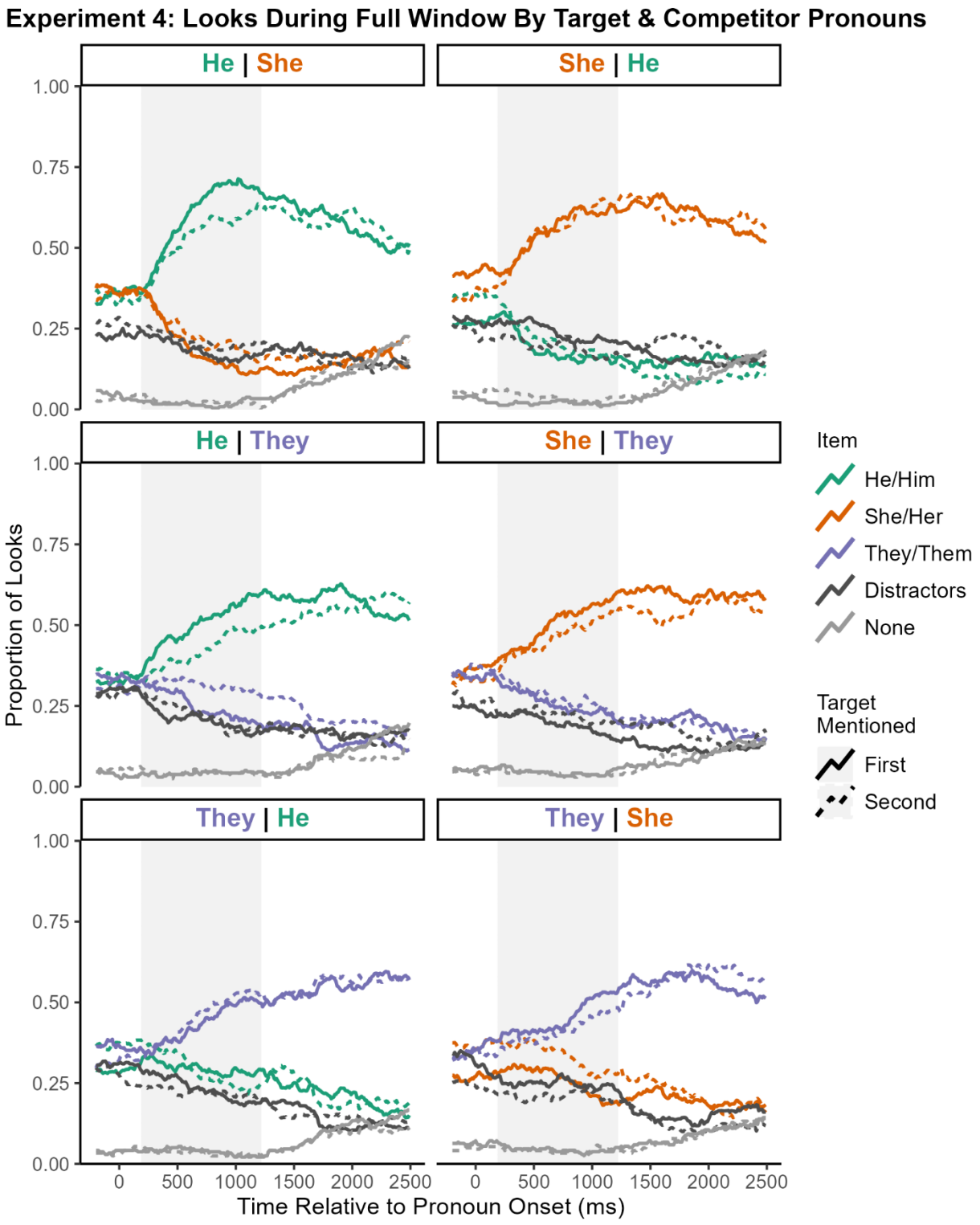


Figure 4.6: Experiment 4: Full Window. Proportions of looks to the target, competitor, distractor (average of 4), and no characters, split by target pronoun, competitor pronoun, and order of mention conditions. The gray box indicates the analysis region, starting 200ms after pronoun onset and ending at 1210ms, the earliest shape word onset across stimuli.

The primary analysis window, shown in grey in Figure 4.6, was offset by 200ms after the onset of the pronoun, the estimated time it takes to plan and execute a saccade in response to the auditory stimulus (Hallett, 1986). It continued until 1210ms after the pronoun onset, which was the earliest shape word onset across all of the stimuli (range = 1216–1790ms, *M* = 1364ms, *SD* = 103ms). Results were analyzed with dynamic generalized mixed-effects models, predicting whether participants looked at the target character (=1) or not (=0) at each time point (Brown-Schmidt et al., 2020; Cho et al., 2018). Observations were down-sampled to 10ms bins, where bins that included >5ms of a fixation on or a saccade to the target (McMurray et al., 2009, 2019) were coded as 1, bins that included <5ms were coded as 0, and bins that included 5ms were coded as 1 if they followed a bin that was coded as 1 and 0 if not. Aside from this, the data was not aggregated across trials or participants. The model included a fixed effect for Trend (timestep during trial, mean-centered) to capture linear changes across the trial in the level of fixations to the target. To account for autocorrelation between time points, the model included an AR(1) term, which captures whether the participant was looking at the target in the prior timestep. To calculate AR(1) for the start of the analysis window, timesteps for 180ms and 190ms were included in the data, and then the first timestep with missing data for AR(1) was excluded prior to estimation, resulting in 103 data points for each trial.

The fixed effect of Pronoun Pair was coded with orthogonal Helmert contrasts, with the first contrast comparing trials with *they* target characters to trials with *he* or *she* target characters (They|HeShe vs HeShe|They + HeShe|SheHe), and the second contrast comparing trials with *he* or *she* target and *they* competitor characters to trials with *he* or *she* target and s*he* or *he* competitor characters (HeShe|They vs HeShe|SheHe). The fixed effect of Order was mean-center effects coded, comparing trials where the target character was mentioned second to trials where the target character was mentioned first. Figure 4.7 shows the proportion of looks to the target and competitor characters during the analysis window, comparing the 3 Pronoun Pair and 2 Order of Mention conditions.

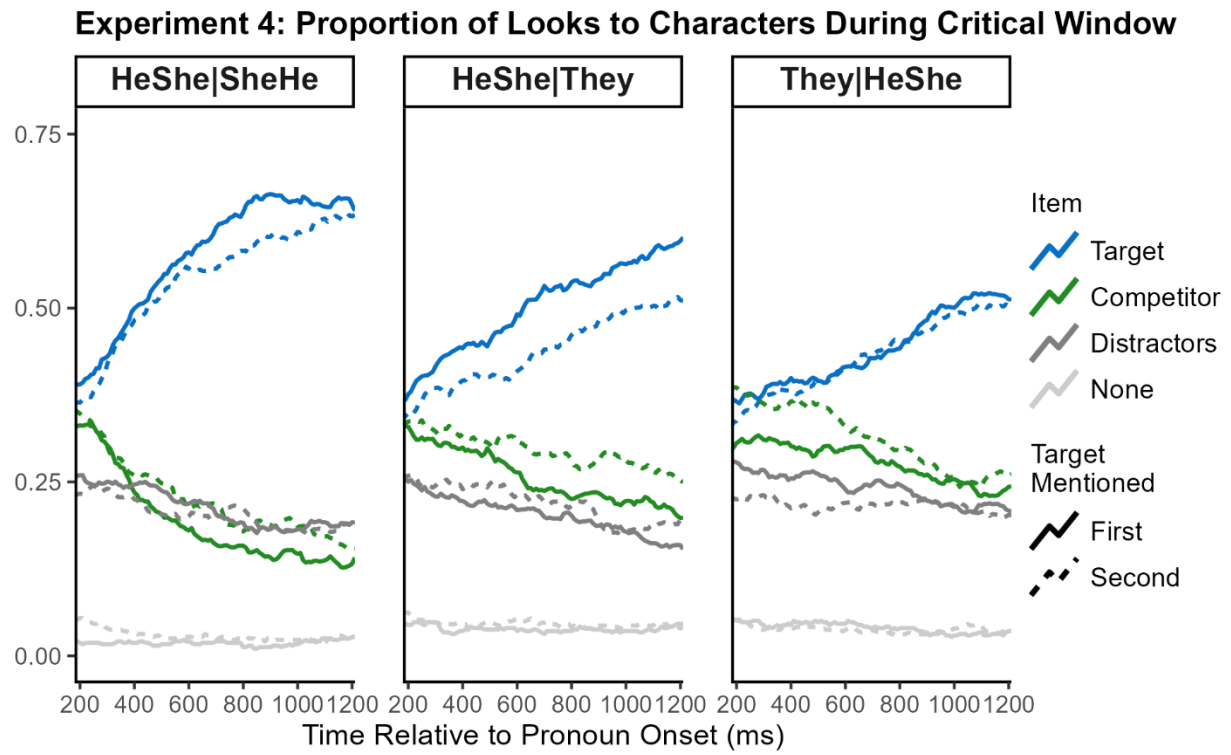


Figure 4.7: Experiment 4: Analysis Window. Proportion of looks to target, competitor, distractor (average of 4), and no characters, comparing between the 3 pronoun pair and 2 order of mention conditions. The window starts 200ms after pronoun onset and ends at 1210ms, the earliest shape word onset across stimuli.

The maximal random effects structure (Baayen et al., 2008; Barr et al., 2013) included by-participant and by-item slopes for Pronoun Pair, Order, AR(1), Trend (time point during trial), and Trial Number (time point during experiment), with items defined as the 60 story frames that named the 2 characters. The *lme4* and *buildmer* packages in R identified the most complex random effects structure that would converge (Bates et al., 2015; R Core Team, 2023; Voeten, 2023), which included by-participant slopes for AR, Order, and Trial Number and by-item slopes for Order and Trend (Table 4.2). The AR(1) effect was significant (β = 9.97, z = 91.37, $p < .001$), reflecting the fact that participants were more likely to be looking at the target during the current timestep if they had been looking at the target during the previous timestep. Trend was not significant

(β = 0.05, z = 1.37, p = .17), indicating that the overall level of target fixations did not significantly increase or decrease in a linear fashion over the course of the trial.

Both contrasts for Pronoun Pair were significant: Participants were more likely to look at the target character after the onset of *he* and *she* than after the onset of *they*, across Order conditions (β = 0.28, z = 6.00, p < .001). After the onset of *he* or *she*, participants were more likely to look at the target character if the competitor character used he/him or she/her than if the competitor character used they/them (β = 0.28, z = 5.16, p < .001). Visual inspection of the data shows that in stories using *he* and *she*, looks to the target diverge from looks to the competitor and reach a proportion of 0.5 in the first quarter of the analysis window. In stories using *they*, looks to the target diverge from looks to the competitor in the first quarter of the analysis window, but do not reach 0.5 until the last quarter (Figure 4.7).

In the primary model, the main effect of Order was not significant, indicating that listeners were not more likely to be looking at target characters mentioned first than target characters mentioned second (β = 0.08, z = 1.06, p = .29). Both interactions of Order with Pronoun Pair were nonsignificant (They|HeShe vs HeShe|They + HeShe|SheHe: β = 0.14, z = 1.52, p = .13; HeShe|They vs HeShe|SheHe: β = -0.18, z = -1.65, p = .10). The lack of an order effect was a surprise, given that this was a robust result in prior research using a similar paradigm (Arnold et al., 2000, 2007; Brown-Schmidt & Toscano, 2017). Post-hoc models explored the order of mention effect in each condition separately (Table A4.6, Table A4.7, Table A4.8). These analyses revealed a significant effect of Order in the HeShe|They condition (β = 0.22, z = 2.20, p < .05), but not in the HeShe|SheHe (β = 0.05, z = 0.47, p = .64) or They|HeShe conditions (β = 0.00, z = -0.03, p = .97).

Table 4.2: Experiment 4: Looks to the Target Character. Model results for the effects of Pronoun Pair, Order, AR(1), and Trend on the likelihood of looking at the target character (=1) or not (=0).

Experiment 4: Looks to the Target Character

	Looks to Target			
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	-4.715	0.084	-56.037	**<0.001**
AR(1) (0, 1)	9.973	0.109	91.367	**<0.001**
Pronoun Pair: They|HeShe (-.66) **vs HeShe|They** (+.33) **+ HeShe|SheHe** (+.33)	0.282	0.047	6.004	**<0.001**
Pronoun Pair: HeShe|They (-.5) **vs HeShe|SheHe** (+.5)	0.277	0.054	5.158	**<0.001**
Order: Target Mentioned Second (-.5) vs First (+.5)	0.084	0.079	1.059	0.290
Trend (mean-centered)	0.052	0.038	1.370	0.171
Pronoun Pair (They|HeShe vs HeShe|They + HeShe|SheHe) * Order	0.143	0.094	1.521	0.128
Pronoun Pair (HeShe|They vs HeShe|SheHe) * Order	-0.177	0.107	-1.651	0.099
Random Effects				
$T_{00\ Story}$	0.007			
$T_{00\ Participant}$	0.196			
$T_{11\ Order\ |\ Story}$	0.024			
$T_{11\ Trend\ |\ Story}$	0.005			
$T_{11\ AR(1)\ |\ Participant}$	0.271			
$T_{11\ Order\ |\ Participant}$	0.112			
$T_{11\ Trial\ Number\ |\ Participant}$	0.025			
$\rho_{01\ Order\ |\ Story}$	0.964			
$\rho_{01\ Trend\ |\ Story}$	0.214			
$\rho_{01\ AR(1)\ |\ Participant}$	-0.299			
$\rho_{01\ Order\ |\ Participant}$	0.091			
$\rho_{01\ Trial\ Number\ |\ Participant}$	0.527			
$N_{Participant}$	30			
N_{Story}	60			
Observations	296022			

After noting differences between the He|She and She|He trials in Figure 4.6, an exploratory analysis tested the effect of Target Pronoun. As in previous experiments, it was coded with orthogonal Helmert contrasts, with the first contrast comparing they/them to he/him + she/her target trials, and the second contrast comparing she/her to he/him target trials. The comparison between *he* and *she* trials was not significant (β = 0.04, z = 0.73, p = .47) (Table A4.9). Additional exploratory analyses tested if Trend and AR(1) interacted with Pronoun Pair and Order, finding no significant effects in addition to those reported in the main model (Table A4.10, Table A4.11).

4.5 Discussion

Experiment 4 tested online processing of singular *they* coreferring with proper names, in a context where listeners can come to anticipate it for certain referents. Participants learned about two he/him, two she/her, and two they/them characters and passed a pretest to ensure that they learned the mappings between the names and images. In the eyetracking trials, each of the six characters was pictured in a colored shape. The stories named two characters—always a pair with different pronouns—and used a verb that allows an upcoming pronoun to refer to either of the named characters (e.g., *Jaime is painting a portrait of Sam, as some paint is spilling on the floor*). The pronoun phrase described the target character's location (e.g., *They are standing in a blue triangle*), leaving about 1200ms between the onset of the pronoun and the onset of the disambiguating shape word to measure how listeners used the pronoun to identify the target character. Trials varied by two within-participants conditions: the pronouns of the two named characters (HeShe|SheHe, HeShe|They, They|HeShe) and whether the pronoun referred to the character mentioned first or second.

Visual inspection of the data suggests that listeners do comprehend *they* as singular and use this information to identify which character is being referred to, just to a lesser degree than with *he* and *she*. Looks to the target diverged from looks to the competitor at about 500ms after the onset of *they*, compared to about 200ms after the onset of *he* or *she.* This is similar to the

patterns seen in young children (Arnold et al., 2007; Song & Fisher, 2005) and in adult second language learners (Cunnings et al., 2017; Grüter et al., 2012; Speyer & Schleef, 2019). Singular *they*, however, did not show an order of mention effect, suggesting that listeners may be using different cues to disambiguate singular *they* from plural *they* than when *he* or *she* could refer to multiple referents (e.g., if both of the named characters used he/him). After each trial, participants judged whether the story matched the scene, and they were not less likely to say that stories used singular *they* matched, but their responses were slower than for *he* and *she*.

Even though the competitor character's pronouns were never used in a given story, they did affect processing. Within *he* and *she* trials, listeners were less likely to be looking at the target character when the competitor character used they/them than when the competitor character used she/her or he/him. One explanation is that participants were paying more attention in general to the they/them characters, since they were atypical or more difficult, or because they were inferred to be the goal of the study. Examining the fixations during the 1000ms where the screen was displayed before the audio started (Figure A4.2) and during the first phrase when the two characters were named (Figure A4.1) rules out this explanation, because participants were not more likely to be looking at the they/them characters than the he/him and she/her characters during either window.

One unexpected result is that the *he* and *she* trials did not entirely replicate the order of mention effect, which has been robust in previous experiments (Arnold et al., 2000, 2007; Brown-Schmidt & Toscano, 2017; Falandays et al., 2020). While there was a clear order of mention effect in the HeShe|They trials, the effect of order was not significant in the HeShe|SheHe trials. Breaking the data down further shows that the expected order pattern appeared for He|She trials, but not She|He Trials. Replication data collection is in progress to determine whether this is a consistent pattern. With a larger data set, it will also be possible to conduct a dynamic tree-based item response analysis (Cho et al., 2020), which can test additional hypotheses by making a distinction between looks to the competitor character and looks to the distractor characters.

CHAPTER 5

General Discussion

Using singular *they* when a referent's gender is unknown or unspecified has become increasingly common and accepted by U.S. English speakers in recent years. Using they/them pronouns is not exclusive to nonbinary people, but it is a particularly common choice for them, since singular *they* is the most common gender-neutral alternative to he/him and she/her (Cassian, 2021, 2022; Cheung et al., 2020; C. Parks et al., 2023). The majority of U.S. adults (50–70% across age groups) know at least a little about people using they/them pronouns, and 11–46% personally know someone who uses they/them in 2021, up from 8–32% in 2018. While these proportions are highest in younger generations, they have been increasing across all generations, as well as across political affiliations and religious groups (Minkin & Brown, 2021; Parker et al., 2019). This increase in familiarity with gender-neutral pronouns corresponds to increases in both the visibility and the size of the trans and gender diverse (TGD) community. Estimates based on probability-sampling surveys suggest that 1 in 200 adults in the U.S. consider themselves[4] transgender (Crissman et al., 2017; Herman et al., 2022; Meerwijk & Sevelius, 2017). This number has been increasing since the mid-2010s (CDC, 2021), especially as population-based surveys are changing their question designs to recognize nonbinary people who fall under the TGD umbrella, but may not also identify as transgender. Among Gen Z adults (ages 18–26 in 2021 and 2022), 2.3% identify as trans and 3.3% as *none of these* (out of *man*, *woman*, *transgender*), which corresponds to the population of the fifth-largest city in the U.S. (Twenge, 2023).

The backlash to an increasing number of teens and young adults openly expressing TGD identities has been severe. 2023 has seen a record number of bills introduced in U.S. state legislatures targeting healthcare for, civil rights protections for, and education about TGD people

(Human Rights Campaign, 2023). Many of these bills include policies preventing people from respecting that someone has changed their pronouns, particularly in educational contexts (e.g., the recently-passed TN S.B. 0466/H.B. 1269, 2023). “Using pronouns” has become a symbol of whether or not you acknowledge and respect TGD people’s existence. Indeed, people describe being misgendered as alienating, devaluing, invalidating, and painful (Cordoba, 2020; Goldberg et al., 2019; Gunn, 2020; A. Johnson, 2019; Pitcher, 2017; Saltzburg & Davis, 2010; Truszczynski et al., 2020), and people who are misgendered more frequently show higher rates of depression, stress, suicidality, disordered eating, body dissatisfaction, and low self-esteem (Galupo et al., 2020; McLemore, 2015, 2018; Mitchell et al., 2021; The Trevor Project, 2020).

However, this opposition represents a loud minority. Many people are open to learning: across age groups, 40–60% of U.S. adults say they would be somewhat or very comfortable using a gender-neutral pronoun for someone who asked (Minkin & Brown, 2021). Still, learning a new language form can be hard, and many people continue to make errors, despite supportive intentions. This often results in seemingly-counterintuitive errors like “she uses they/them pronouns” (e.g., Towers, 2021 in §0.1). I began by arguing that errors like these occur because learning to use they/them pronouns requires a change in pronoun production processes. Instead of selecting pronouns based on morphosyntactic gender features of a name (e.g., Schmitt et al., 1999) or based on semantic/conceptual gender features of a person (e.g., Antón-Méndez, 2010), speakers may need to retrieve an episodic memory of a person’s stated pronouns or the pronouns they have observed other speakers using in reference to that person. To that end, this dissertation investigated how people learn to associate pronouns with a person, how they use that information in language production and comprehension, and what factors can support learning.

Experiment 1 demonstrated that speakers can learn to associate pronouns with a person, in a context where the use of they/them could not be predicted from the name. While explicit memory for a person’s pronouns strongly predicted accurate production, it did not guarantee it. In Experiment 2, reading a brief PSA about gendered language increased the proportion of speakers

who produced singular *they* at least once and their accuracy when doing so—arguably by motivating participants to pay more attention to information about the character's pronouns during the introductions, then to attempt to retrieve the information later during the sentence completion task. Experiment 3 then showed that when information about a person's pronouns is made more salient by directly stating it during introductions and made more available by including it on nametags, speakers can and do use this information to correctly produce singular *they*. These results all support a model where speakers can use episodic memory for a person's stated pronouns to select which pronoun to produce, behavior that is not accounted for in current processing models.

One of the primary focuses of research on pronoun production has been *when* speakers choose to use pronouns instead of other referring expressions (Arnold & Zerkle, 2019), but the subsequent choice of *which* pronoun to use is less well understood. If applied to proper names and pronouns referring to people, production models either assume that speakers select pronouns based on morphosyntactic features of the name, or based on semantic/conceptual features of a person. As an example of the first approach, Schmitt et al.'s (1999) model of lexical access in pronoun production argues that if a lexical concept is activated and sufficiently "in focus" in the discourse, the speaker will produce a pronoun instead of a full noun phrase. In this model, tested with object labels in German, activating the lexical concept for the noun also activates the corresponding grammatical gender node (masculine, feminine, or neuter). If the speaker uses a pronoun, the gender node is selected in order to produce the correct pronoun (Jescheniak & Levelt, 1994; Levelt et al., 1999; Roelofs, 1992). When producing gender-marked pronouns or determiners, competition between different forms can arise from the grammatical gender features of other lexical concepts that are also activated (Schiller & Caramazza, 2003). If each lexical concept can only correctly be associated with a single grammatical gender, models of grammatical gender selection may not need to include competition between multiple grammatical gender features activated by the same lexical concept. The closest analogue may be languages

where the singular and plural forms of determiners vary for the same grammatical gender, and one approach argues that the singular form is activated by default and can interfere with activating the plural form (Jescheniak et al., 2014; Schriefers et al., 2002).

In models of grammatical gender processing, it is reasonable to assume that once a speaker decides to produce a pronoun, there is typically only one correct pronoun to use. If extended from pronouns referring to objects to third person gendered pronouns referring to people, Schmitt et al.'s model might predict that the lexical entry for a masculine name would automatically activate the masculine gender node, which is linked to he/him pronouns, and the lexical entry for a feminine name would automatically activate the feminine gender node, which is linked to she/her pronouns. However, this approach would need to be modified to account for how speakers use singular *they*. One option would be to add a neutral gender node,[13] which is linked to singular they/them pronouns. The lexical entry for a name could activate the neutral gender node, in addition to or instead of the masculine or feminine nodes. This is similar to how Konnelly & Cowper's morphosyntax model of gender agreement (described in §0.3) could function in a processing model. They propose that "innovative" speakers, who only accept specific definite antecedents for singular *they* that are gender-unspecified, can accommodate a proper name by adding a new lexical entry for the name that does not have a masculine or feminine feature (2020).

This gets us closer to accounting for the use of singular *they* in the current experiments by allowing a name to activate multiple gender nodes. In Experiments 1 & 2, the masculine names would activate the masculine gender node and possibly also the neutral gender node, and the feminine names would activate the feminine gender node and possibly also the neutral gender node. If the neutral gender node is activated, it is presumably in competition with the masculine or feminine node, with speakers trying to use *they* instead of defaulting to *he* or *she*. However, it

[13] *A related question is if the model would need separate nodes for gender-unspecified and gender-specified singular* they*. But since the experiments here only elicit gender-specified* they*, I focus on that form.*

is clear that speakers are using more than the lexical information for proper names to select pronouns. None of the current experiments tested a context where names provided enough information to predict pronouns, such as a set of characters where everyone with masculine names used he/him, everyone with feminine names used she/her, and everyone with gender-neutral names used they/them. Instead, Experiments 1 & 2 only used masculine and feminine names, with one third of the characters using they/them and the other two thirds using the expected he/him or she/her. Experiment 3 used all gender-neutral names, where the use of he/him, she/her, or they/them could not be predicted from the name. In both contexts, using singular *they* required information in addition to the lexical features or knowledge about gender distributions associated with the name. Participants had to associate they/them with the particular character, i.e., *this* Alex, not all Alex's in general.

Instead of proposing that pronouns are selected via lexical information associated with the name, pronoun production models can assume that speakers select pronouns based on semantic or conceptual information about the person's gender (Ackerman, 2019). Here, psycholinguistics research has mostly focused on errors in adult second language learning. For example, English possessive pronouns agree with the possessor (e.g., in *Jaime gave the apple to their brother*, the pronoun agrees with *Jaime*), which can result in errors by native speakers of languages where possessive pronouns agree with the possessee (e.g., the pronoun would agree with *brother* instead of *Jaime*) (Pozzan & Antón-Méndez, 2017). This is more of an issue about which person the pronoun agrees with, not which gendered form should be selected for the person. Other experiments have looked at errors by native speakers of languages that do not mark gender in third-person singular pronouns, such as Finnish and spoken Mandarin Chinese (Pozzan & Antón-Méndez, 2017; Siewierska, 2013). This results in errors between he/him and she/her, where it is assumed that the speaker has the correct inference about the person's gender, but not a

consistent mapping between gender categories and which pronouns to produce[14] (Ackerman, 2019), since it is not information their native language requires them to encode (Slobin, 1996).

An approach where speakers select pronouns based on semantic or conceptual information about the person's gender can better account for they/them pronouns, particularly if you assume that information about a person's pronouns could be encoded as part of conceptual gender. Traditional approaches treat conceptual gender as primarily automatic inferences (see the mismatching gender stereotype experiments in §0.4.1), but one option is to assume episodic information can be added. A person stating that they use they/them pronouns or other speakers are using they/them pronouns for that person could prompt a revision of the original gender inference that directed to he/him or she/her.[15] This model would also need to be modified to account for multiple pronouns being associated with a person, instead of having a one conceptual gender inference to one set of pronouns mapping. Speakers often alternate between they/them and he/him or she/her, like the errors in these experiments or in the Demi Lovato article (Towers, 2021). Alternating between pronouns can also be intentional when referring to people who accept multiple pronouns (e.g., Gurley, 2021; Monáe & Lore, 2022). A one-to-one model also does not account for gender-unspecified specific singular *they*, particularly when it is used when the speaker *does* know the referent's gender, but does not include that information. In this case, the conceptual gender information is not driving pronoun selection, because the speaker is choosing

---

[14] *Accounts of grammatical gender errors assume the opposite. Grammatical gender marking errors remain common even in highly-proficient learners, particularly if they are native speakers of a language without grammatical gender (e.g., English speakers learning Spanish) (Antón-Méndez, 2010; Grüter et al., 2012). These errors occur when a speaker selects the wrong grammatical gender feature for a word, then produces the pronoun corresponding to that wrong grammatical gender category.*

[15] *Here, I can only claim that information about gendered language forms is added to knowledge about a person, which could but does not necessarily correspond to changing an underlying gender categorization. See Friedman (2014) for a theoretical model and behavioral data about how people revise their original automatic categorizations of a person's gender.*

to not specify gender. Overall, the behavior of participants producing singular *they* in the first three experiments demonstrates that speakers are incorporating additional sources of information and choosing between more options than are accounted for by existing processing models, where pronouns are selected based on the lexical information connected to the name or the gender categorization connected to the person.

Experiment 4 turned from production to comprehension, but with the same underlying question: specific definite singular *they* does show a processing cost compared to *he* and *she*, but it is still unclear how large and how persistent those costs are, and how singular *they* is fitting into the existing processing mechanisms comprehenders have for *he* and *she*. Results from a visual world study showed that, in a context where participants could come to expect certain characters to be referred to with they/them, listeners were able to use singular *they* to identify the referent, just to a lesser degree than with *he* and *she.* This resembles patterns seen in young children (Arnold et al., 2007; Song & Fisher, 2005) and in adult second language learners (Cunnings et al., 2017; Grüter et al., 2012; Speyer & Schleef, 2019). One open question is how these findings will replicate in future years as people continue to gain experience with using they/them pronouns, and if the learning trajectory will be similar to those observed for other pronouns during language development and second language acquisition.

Unlike prior findings for when *he* or *she* is ambiguous between two referents (Arnold et al., 2000, 2007; Brown-Schmidt & Toscano, 2017; Falandays et al., 2020) and preliminary findings for singular *they* using a different task (Arnold et al., 2023), participants in Experiment 4 did not show an order of mention effect for singular *they*. The pronouns in these stories can refer to either of the two named characters, but is more likely to refer to the character named first than the character named second (Gernsbacher, 1989; Kaiser & Trueswell, 2011). Listeners can use this information to help disambiguate the referent, and thus start looking at the target character—which the visual world paradigm takes to mean that they understand who the sentence is referring to—earlier for characters mentioned first. This pattern was not observed in *they* trials. The third

finding in Experiment 4 was that after hearing *he* or *she*, listeners were more likely to be looking at the target character if the competitor character used she/her or he/him than if the competitor character used they/them. The lack of an order of mention effect and the observation that they/them characters behaved differently as competitors than he/him and she/her characters both suggest that while listeners can successfully comprehend *they* as singular, they may be using different strategies to do so, leaving a number of follow-up research questions.

I want to emphasize here that the processing cost for singular *they* is large enough to be statistically significant and relevant for understanding language processing mechanisms, but that this does not necessarily correspond to meaningful differences in processing fluency—people's subjective experience of how easy it was to understand the story (Alter & Oppenheimer, 2009). In this experiment, visual inspection of the data shows that around 50% of participants were looking at the target character half a second after the onset of *he* or *she*, compared to around 35% half a second after the onset of *they* (Figure 4.7). Similarly, in one reading time study, indefinite gender-unspecified *they* (e.g., *a person, someone*) took about 100ms longer than *he* and *she*, which the authors note is a smaller, more easily accommodated processing cost than other types of gender and number agreement mismatches (Sanford & Filik, 2007, see §0.4.2). The processing fluency account predicts that language that is more difficult to understand elicits more negative attitudes about the speaker (Dragojevic, 2020; Dragojevic et al., 2017; Dragojevic & Giles, 2016, see §4.1). However, it is an open question how processing costs in psycholinguistics measures correspond to processing fluency—how the time course in a visual world study matches how hard a listener consciously perceives singular *they.* Like Sanford & Filik I would argue that, when compared to the processing costs seen in other experiments, the processing costs observed here with singular *they* are within the range that comprehenders can normally accommodate. Instead of singular *they* being dramatically or permanently more difficult to understand, reference resolution is already complicated and ambiguous, even in contexts that are less socio-politically marked.

From a theoretical perspective, the four experiments in this dissertation demonstrate how singular *they* pushes language processing models to incorporate additional factors—especially social ones—to account for the full range of people's language use. From a practical perspective, these experiments show that while learning singular *they* is hard, it *is* doable. Multiple strategies can support learning, including providing information about why paying attention to gendered language is important, showing examples of singular *they* in use, making information about pronouns more salient by including it in introductions, and making information about pronouns more accessible by including it in places like nametags and display names. Learning to use they/them pronouns is worthwhile, because misgendering is a key aspect of how respected and welcome trans and gender diverse people feel in a community, and reducing misgendering is something that individual allies have the opportunity to concretely change.

REFERENCES

Abbott, B. (2006). Definite and indefinite. In *Encyclopedia of language & linguistics* (pp. 392–399). Elsevier. https://doi.org/10.1016/B0-08-044854-2/01089-0

Ackerman, L. (2018a). *Processing singular* they *with generic and specific antecedents* [Poster]. 24th Architectures and Mechanisms of Language Processing. https://osf.io/ztrsy/

Ackerman, L. (2018b). *Toward an understanding of how nonbinary pronouns are learned and processed* [Poster]. Socially Situated Language Processing. https://osf.io/zacp7/

Ackerman, L. (2019). Syntactic and cognitive issues in investigating gendered coreference. *Glossa: A Journal of General Linguistics*, *4*(1), 117. https://doi.org/10.5334/gjgl.721

Ackerman, L. (2020). *Social exposure to gender-variance influences the real-time processing of pronouns* [Poster]. 33rd Annual CUNY Human Sentence Processing Conference. https://osf.io/98mhp/

Ackerman, L., Riches, N., & Wallenberg, J. (2017). *The ambiguity of natural gender in coreference dependency formation* [Poster]. 23rd Architectures and Mechanisms of Language Processing. https://osf.io/vwj8p/

Adamsky, C. (1981). Changes in pronominal usage in a classroom situation. *Psychology of Women Quarterly*, *5*(5), 773–779. https://doi.org/10.1111/j.1471-6402.1981.tb01101.x

Ahn, B., & Conrod, K. (2022). *Three ways to rate* themself [Conference talk]. 35th Annual Conference on Human Sentence Processing.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(38), 419–439. https://doi.org/10.1006/jmla.1997.2558

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, *13*(3), 219–235. https://doi.org/10.1177/1088868309341564

American Civil Liberties Union. (2023). *Mapping attacks on LGBTQ rights in U.S. state legislatures*. https://web.archive.org/web/20230529140006/https:/www.aclu.org/legislative-attacks-on-lgbtq-rights

Ansara, Y. G., & Hegarty, P. (2014). Methodologies of misgendering: Recommendations for reducing cisgenderism in psychological research. *Feminism & Psychology*, *24*(2), 259–270. https://doi.org/10.1177/0959353514526217

Antón-Méndez, I. (2010). Gender bender: Gender errors in L2 pronoun production. *Journal of Psycholinguistic Research*, *39*(2), 119–139. https://doi.org/10.1007/s10936-009-9129-z

Antón-Méndez, I., Nicol, J. L., & Garrett, M. F. (2002). The relation between gender and number agreement processing. *Syntax*, *5*(1), 1–25. https://doi.org/10.1111/1467-9612.00045

Arechar, A. A., & Rand, D. G. (2021). Turking in the time of COVID. *Behavior Research Methods*, *53*(6), 2591–2595. https://doi.org/10.3758/s13428-021-01588-4

Ariel, M. (2006). Accessibility theory. In *Encyclopedia of language & linguistics* (pp. 15–18). Elsevier. https://doi.org/10.1016/B0-08-044854-2/04291-7

Arnold, J. E., Brown-Schmidt, S., & Trueswell, J. (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Processes*, *22*(4), 527–565. https://doi.org/10.1080/01690960600845950

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition*, *76*(1), B13–B26. https://doi.org/10.1016/S0010-0277(00)00073-1

Arnold, J. E., Marquez, A., Li, J., & Franck, G. (2022). Does nonbinary *they* inherit the binary pronoun production system? *Glossa Psycholinguistics*, *1*(1). https://doi.org/10.5070/G601183

Arnold, J. E., Mayo, H. C., & Dong, L. (2021). My pronouns are they/them: Talking about pronouns changes how pronouns are understood. *Psychonomic Bulletin & Review*, *28*, 1688–1697. https://doi.org/10.3758/s13423-021-01905-0

Arnold, J. E., Payst, N., Kakoki, E., & Wang, L. (2023). *Online comprehension of nonbinary* they *is easier with supporting linguistic context* [Poster]. 36th Annual Conference on Human Sentence Processing.

Arnold, J. E., Venkatesh, R., Vig, Z. A., & Scott, A. (2022). *University students produce nonbinary singular* they *almost as well as* he *or* she [Conference talk]. 35th Annual Conference on Human Sentence Processing.

Arnold, J. E., & Zerkle, S. A. (2019). Why do people produce pronouns? Pragmatic selection vs. rational models. *Language, Cognition and Neuroscience*, *34*(9), 1152–1175. https://doi.org/10.1080/23273798.2019.1636103

Austen, J. (1815). *Emma*.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baldwin, A., Dodge, B., Schick, V. R., Light, B., Scharrs, P. W., Herbenick, D., & Fortenberry, J. D. (2018). Transgender and genderqueer individuals' experiences with health care providers: What's working, what's not, and where do we go from here? *Journal of Health Care for the Poor and Underserved*, *29*(4), 1300–1318. https://doi.org/10.1353/hpu.2018.0097

Balhorn, M. (2004). The rise of epicene *they*. *Journal of English Linguistics*, *32*(2), 79–104. https://doi.org/10.1177/0075424204265824

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

Barbee, H., & Schrock, D. (2019). Un/gendering social selves: How nonbinary people navigate and experience a binarily gendered world. *Sociological Forum*, *34*(3), 572–593. https://doi.org/10.1111/socf.12517

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Baskervill, W. M., & Sewell, J. W. (1895). *An English grammar*. Project Gutenburg.

Bates, D. M., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baum, J., Brill, S., Brown, J., Delpercio, A., Kahn, E., Kenney, L., & Nicoll, A. (2012). *Supporting and caring for our gender-expansive youth: Lessons from the Human Rights Campaign's youth survey*. Human Rights Campaign Foundation & Gender Spectrum. https://web.archive.org/web/20230323072407/https://assets2.hrc.org/files/assets/resources/Gender-expansive-youth-report-final.pdf

Bjorkman, B. M. (2017). Singular *they* and the syntactic representation of gender in English. *Glossa: A Journal of General Linguistics*, *2*(1), 80. https://doi.org/10.5334/gjgl.374

Blaubergs, M. S. (1980). An analysis of classic arguments against changing sexist language. *Women's Studies International Quarterly*, *3*(2–3), 135–147. https://doi.org/10.1016/S0148-0685(80)92071-0

Bockting, W. O., Miner, M. H., Swinburne Romine, R. E., Hamilton, A., & Coleman, E. (2013). Stigma, mental health, and resilience on an online sample of the US transgender population. *American Journal of Public Health*, *103*(5), 943–951. https://doi.org/10.2105/AJPH.2013.301241

Bodine, A. (1975). Androcentrism in prescriptive grammar: Singular *they*, sex-indefinite *he*, and *he or she*. *Language in Society*, *4*(2), 129–146. https://doi.org/10.1017/S0047404500004607

Bradley, E. D. (2020). The influence of linguistic and social attitudes on grammaticality judgments of singular *they*. *Language Sciences*, *78*, 101272. https://doi.org/10.1016/j.langsci.2020.101272

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usages*. Cambridge University Press.

Brown-Schmidt, S., Naveiras, M., De Boeck, P., & Cho, S.-J. (2020). Statistical modeling of intensive categorical time-series eye-tracking data using dynamic generalized linear mixed-effect models with crossed random effects. In *Psychology of learning and motivation* (Vol. 73, pp. 1–31). Elsevier. https://doi.org/10.1016/bs.plm.2020.06.004

Brown-Schmidt, S., & Toscano, J. C. (2017). Gradient acoustic information induces long-lasting referential uncertainty in short discourses. *Language, Cognition and Neuroscience*, *32*(10), 1211–1228. https://doi.org/10.1080/23273798.2017.1325508

Buchanan, N. T., Perez, M., Prinstein, M. J., & Thurston, I. B. (2021). Upending racism in psychological science: Strategies to change how science is conducted, reported, reviewed, and disseminated. *American Psychologist*, *76*(7), 1097–1112. https://doi.org/10.1037/amp0000905

Budge, S. L., Rossman, H. K., & Howard, K. A. S. (2014). Coping and psychological distress among genderqueer individuals: The moderating effect of social support. *Journal of LGBT Issues in Counseling*, *8*(1), 95–117. https://doi.org/10.1080/15538605.2014.853641

Burgwal, A., Gvianishvili, N., Hård, V., Kata, J., Garcia Nieto, I., Orre, C., Smiley, A., Vidic, J., & Motmans, J. (2019). Health disparities between binary and non-binary trans people: A community-driven survey. *International Journal of Transgenderism*, *20*(2–3), 218–229. https://doi.org/10.1080/15532739.2019.1629370

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review*, *106*(4), 676–713. https://doi.org/10.1037/ 0033-295X.106.4.676

Butler, C., Joiner, R., Bradley, R., Bowles, M., Bowes, A., Russell, C., & Roberts, V. (2019). Self-harm prevalence and ideation in a community sample of cis, trans and other youth. *International Journal of Transgenderism*, *20*(4), 447–458. https://doi.org/10.1080/15532739.2019.1614130

Cacciari, C., & Padovani, R. (2007). Further evidence of gender stereotype priming in language: Semantic facilitation and inhibition in Italian role nouns. *Applied Psycholinguistics*, *28*(2), 277–293. https://doi.org/10.1017/s0142716407070142

Cameron, J. J., & Stinson, D. A. (2019). Gender (mis)measurement: Guidelines for respecting gender diversity in psychological research. *Social and Personality Psychology Compass*, *13*(11), e12506. https://doi.org/10.1111/spc3.12506

Camilliere, S., Izes, A., Leventhal, O., & Grodner, D. J. (2021). *They* is changing: Pragmatic and grammatical factors that license singular *they*. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*.

Carlson, G. (2006). Generic reference. In *Encyclopedia of language & linguistics* (pp. 14–18). Elsevier. https://doi.org/10.1016/B0-08-044854-2/00274-1

Carpenter, P. A., & Just, M. A. (1983). What your eyes do while your mind is reading. In *Eye movements in reading* (pp. 275–307). Elsevier. https://doi.org/10.1016/B978-0-12-583680-7.50022-9

Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *The Quarterly Journal of Experimental Psychology*, *49A*(3), 639–664. https://doi.org/10.1080/713755647

Cassian. (2021). *Gender census 2021: Worldwide report* [Data Set]. https://web.archive.org/web/20220828234326/https:/www.gendercensus.com/results/2021-worldwide/

Cassian. (2022). *Gender census 2022: Worldwide report* [Data Set]. https://web.archive.org/web/20220827231434/https:/www.gendercensus.com/results/2022-worldwide/

Centers for Disease Control and Prevention. (2014). *2014 Behavioral risk factor surveillance system survey* [Data Set]. https://www.cdc.gov/brfss/annual_data/annual_2014.html

Centers for Disease Control and Prevention. (2015). *2015 Behavioral risk factor surveillance system survey* [Data Set]. https://www.cdc.gov/brfss/annual_data/annual_2015.html

Centers for Disease Control and Prevention. (2019). *2019 Combined youth risk behavior surveillance system high school datasets* [Data Set]. https://www.cdc.gov/healthyyouth/data/yrbs/data.htm

Centers for Disease Control and Prevention. (2021). *2021 Behavioral risk factor surveillance system survey* [Data Set]. https://www.cdc.gov/brfss/annual_data/annual_2021.html

Chandler, J., Rosenzweig, C., Moss, A. J., Robinson, J., & Litman, L. (2019). Online panels in social science research: Expanding sampling methods beyond Mechanical Turk. *Behavior Research Methods*, *51*(5), 2022–2038. https://doi.org/10.3758/s13428-019-01273-7

Chang, T. K., & Chung, Y. B. (2015). Transgender microaggressions: Complexity of the heterogeneity of transgender identities. *Journal of LGBT Issues in Counseling*, *9*(3), 217–234. https://doi.org/10.1080/15538605.2015.1068146

Chen, P., Leventhal, O., Camilliere, S., Blain, K., Izes, A., & Grodner, D. (2023). *Singular* they *in transition: ERP evidence* [Poster]. 36th Annual Conference on Human Sentence Processing.

Cheung, A. S., Leemaqz, S. Y., Wong, J. W. P., Chew, D., Ooi, O., Cundill, P., Silberstein, N., Locke, P., Zwickl, S., Grayson, R., Zajac, J. D., & Pang, K. C. (2020). Non-binary and binary gender identity in Australian trans and gender diverse individuals. *Archives of Sexual Behavior*, *49*(7), 2673–2681. https://doi.org/10.1007/s10508-020-01689-9

Cho, S.-J., Brown-Schmidt, S., Boeck, P. D., & Shen, J. (2020). Modeling intensive polytomous time-series eye-tracking data: A dynamic tree-based item response model. *Psychometrika*, *85*(1), 154–184. https://doi.org/10.1007/s11336-020-09694-6

Cho, S.-J., Brown-Schmidt, S., & Lee, W. (2018). Autoregressive generalized linear mixed effect models with crossed random effects: An application to intensive binary time series eye-tracking data. *Psychometrika*, *83*(3), 751–771. https://doi.org/10.1007/s11336-018-9604-2

Chodzen, G., Hidalgo, M. A., Chen, D., & Garofalo, R. (2019). Minority stress factors associated with depression and anxiety among transgender and gender-nonconforming youth. *Journal of Adolescent Health*, *64*(4), 467–471. https://doi.org/10.1016/j.jadohealth.2018.07.006

Clark, B. A., Veale, J. F., Townsend, M., Frohard-Dourlent, H., & Saewyc, E. (2018). Non-binary youth: Access to gender-affirming primary health care. *International Journal of Transgenderism*, *19*(2), 158–169. https://doi.org/10.1080/15532739.2017.1394954

Conrod, K. (2019a). Names before pronouns: Variation in pronominal reference and gender. *Proceedings of the Northwest Linguistics Conference*.

Conrod, K. (2019b). *Pronouns raising and emerging* (Publication No. 22621294) [Doctoral dissertation, University of Washington]. ProQuest Dissertations & Theses Global.

Corbett, G. G. (2013a). Number of genders. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology. https://web.archive.org/web/20230319221121/https://wals.info/chapter/30

Corbett, G. G. (2013b). Sex-based and non-sex-based gender systems. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology. https://web.archive.org/web/20221203192216/https:/wals.info/chapter/31

Cordoba, S. (2020). *Exploring non-binary genders: Language and identity* (Publication No. 29424421) [Doctoral dissertation, De Montfort University]. ProQuest Dissertations & Theses Global.

Cornish, F. (2006). Discourse anaphora. In *Encyclopedia of language & linguistics* (pp. 631–638). Elsevier. https://doi.org/10.1016/B0-08-044854-2/04290-5

Crissman, H. P., Berger, M. B., Graham, L. F., & Dalton, V. K. (2017). Transgender demographics: A household probability sample of US adults, 2014. *AJPH Transgender Health*, *107*(2), 213–215. https://doi.org/10.2105/AJPH.2016.303571

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*(11), 671–684. https://doi.org/10.1037/h0043943

Cronin, C., & Jreisat, S. (1995). Effects of modeling on the use of nonsexist language among high school freshpersons and seniors. *Sex Roles*, *33*(11–12), 819–830. https://doi.org/10.1007/BF01544781

Cruz, T. M. (2014). Assessing access to care for transgender and gender nonconforming people: A consideration of diversity in combating discrimination. *Social Science & Medicine*, *110*, 65–73. https://doi.org/10.1016/j.socscimed.2014.03.032

Cunnings, I., Fotiadou, G., & Tsimpli, I. (2017). Anaphora resolution and reanalysis during L2 sentence processing. *Studies in Second Language Acquisition*, *39*(4), 621–652. https://doi.org/10.1017/S0272263116000292

DeDe, G., & Knilans, J. (2016). Language comprehension in aging. In H. H. Wright (Ed.), *Cognition, language and aging* (pp. 107–133). John Benjamins Publishing Company. https://doi.org/10.1075/z.200.05ded

Doherty, A., & Conklin, K. (2017). How gender-expectancy affects the processing of *them*. *The Quarterly Journal of Experimental Psychology*, *70*(4), 718–735. https://doi.org/10.1080/17470218.2016.1154582

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLOS ONE*, *18*(3), e0279720. https://doi.org/10.1371/journal.pone.0279720

Dragojevic, M. (2020). Extending the fluency principle: Factors that increase listeners' processing fluency positively bias their language attitudes. *Communication Monographs*, *87*(2), 158–178. https://doi.org/10.1080/03637751.2019.1663543

Dragojevic, M., & Giles, H. (2016). I don't like you because you're hard to understand: The role of processing fluency in the language attitudes process. *Human Communication Research*, *42*(3), 396–420. https://doi.org/10.1111/hcre.12079

Dragojevic, M., Giles, H., Beck, A.-C., & Tatum, N. T. (2017). The fluency principle: Why foreign accent strength negatively biases language attitudes. *Communication Monographs*, *84*(3), 385–405. https://doi.org/10.1080/03637751.2017.1322213

Drucker, Z. (2019). *The gender spectrum collection: Stock photos beyond the binary* [Photographs]. Vice. https://web.archive.org/web/20210209092233/https:/genderphotos.vice.com/

Du Bois, J. W. (2007). The stance triangle. In R. Englebretson (Ed.), *Stancetaking in discourse: Subjectivity, evaluation, interaction* (pp. 139–182). https://doi.org/10.1075/pbns.164.07du

Duffy, S. A., & Keir, J. A. (2004). Violating stereotypes: Eye movements and comprehension processes when text conflicts with world knowledge. *Memory & Cognition*, *32*(4), 551–559. https://doi.org/10.3758/BF03195846

Ehrlich, S., & King, R. (1992). Gender-based language reform and the social construction of meaning. *Discourse & Society*, *3*(2), 151–166.

Eisenberg, M. E., Gower, A. L., McMorris, B. J., Rider, G. N., Shea, G., & Coleman, E. (2017). Risk and protective factors in the lives of transgender/gender nonconforming adolescents. *Journal of Adolescent Health*, *61*(4), 521–526. https://doi.org/10.1016/j.jadohealth.2017.04.014

Fagot, B. I., & Leinbach, M. D. (1993). Gender-role development in young children: From discrimination to labeling. *Developmental Review*, *13*(2), 205–224. https://doi.org/10.1006/drev.1993.1009

Falandays, J. B., Brown-Schmidt, S., & Toscano, J. C. (2020). Long-lasting gradient activation of referents during spoken language processing. *Journal of Memory and Language*, *112*, 104088. https://doi.org/10.1016/j.jml.2020.104088

Farkas, D. (2006). Specificity. In *Encyclopedia of language & linguistics* (pp. 633–635). Elsevier. https://doi.org/10.1016/B0-08-044854-2/04332-7

Flanagan, A. M., & Todd-Mancillas, W. R. (1982). Teaching inclusive generic pronoun usage: The effectiveness of an authority innovation-decision approach versus an optional innovation-decision approach. *Communication Education*, *31*(4), 275–284. https://doi.org/10.1080/03634528209384696

Flowers, A. (2015). *Unisex names data* [Data Set]. FiveThirtyEight. https://github.com/fivethirtyeight/data/tree/master/unisex-names

Fodor, J. D., & Sag, I. A. (1982). Referential and quantificational indefinites. *Linguistics and Philosophy*, *5*(3), 355–398. https://doi.org/10.1007/BF00351459

Foertsch, J., & Gernsbacher, M. A. (1997). Is singular *they* a cognitively efficient substitute for generic *he*? *Psychological Science*, *8*(2), 106–111. https://doi.org/10.1111/j.1467-9280.1997.tb00691.x

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*(2), 78–84. https://doi.org/10.1016/S1364-6613(00)01839-8

Friedman, E. J. (2014). *Cisgenderism in gender attributions: The ways in which social, cognitive, and individuals factors predict misgendering* (Publication No. 3623701) [Doctoral dissertation, The City University of New York]. ProQuest Dissertations & Theses Global.

Galupo, M. P., Pulice-Farrow, L., & Lindley, L. (2020). "Every time I get gendered male, I feel a pain in my chest": Understanding the social context for gender dysphoria. *Stigma and Health*, *5*(2), 199–208. https://doi.org/10.1037/sah0000189

Garnham, A. (2001). A mental models theory of understanding anaphora. In *Mental models and the interpretation of anaphora*. Taylor & Francis.

Garnham, A., & Cowles, H. W. (2006). Reference: Psycholinguistic approach. In *Encyclopedia of language & linguistics* (pp. 427–433). Elsevier. https://doi.org/10.1016/B0-08-044854-2/00801-4

Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition*, *32*(2), 99–156. https://doi.org/10.1016/0010-0277(89)90001-2

GLAAD. (2020). *GLAAD media reference guide—transgender*. GLAAD. https://web.archive.org/web/20200522040917/https://www.glaad.org/reference/transgender

GLSEN. (2020). *Pronouns: A guide from GLSEN*. http://web.archive.org/web/20200423153329/https://www.glsen.org/activity/pronouns-guide-glsen

Goldberg, A. E., Kuvalanka, K., & Dickey, L. (2019). Transgender graduate students' experiences in higher education: A mixed-methods exploratory study. *Journal of Diversity in Higher Education*, *12*(1), 38–51. http://doi.org/10.1037/dhe0000074

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, *17*(3), 311–347. https://doi.org/10.1207/s15516709cog1703_1

Grant, J. M., Mottet, L. A., Tanis, J., Harrison, J., Herman, J. L., & Keisling, M. (2011). *Injustice at every turn: A report of the National Transgender Discrimination Survey*. National Center for Transgender Equality & National Gay and Lesbian Task Force. https://web.archive.org/web/20230127030240/https://transequality.org/issues/resources/national-transgender-discrimination-survey-full-report

Green, P., & MacLeod, C. J. (2016). simr: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Greene, S. B., McKoon, G., & Ratcliff, R. (1992). Pronoun resolution and discourse models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 266–283. https://doi.org/10.1037/0278-7393.18.2.266

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Brill. https://doi.org/10.1163/9789004368811_003

Grosz, B. J., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, *12*(3), 175–204.

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, *28*(2), 191–215. https://doi.org/10.1177/0267658312437990

Gunn, H. A. (2020). *The experience of misgendering among trans and gender diverse people* (Publication No. 2447222123) [Doctoral dissertation, University of Massachusetts Boston]. ProQuest Dissertations & Theses Global.

Gurley, A. (2021, August 12). *"The Crown" star Emma Corrin got candid about their "scary" decision to come out as queer*. BuzzFeed. https://web.archive.org/web/20210813041253/https://www.buzzfeed.com/alexgurley/emma-corrin-queer-journey%0A

Gygax, P., Elmiger, D., Zufferey, S., Garnham, A., Sczesny, S., Von Stockhausen, L., Braun, F., & Oakhill, J. (2019). A language index of grammatical gender dimensions to study the impact of grammatical gender on the way we perceive women and men. *Frontiers in Psychology*, *10*, 1604. https://doi.org/10.3389/fpsyg.2019.01604

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, *8*(4), 439–483. https://doi.org/10.1080/01690969308407585

Hallett, P. E. (1986). Eye movement. In K. Buff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 1, p. 10.1-10.112). Wiley.

Harrison, J., Grant, J., & Herman, J. L. (2012). A gender not listed here: Genderqueers, gender rebels, and otherwise in the National Transgender Discrimination Survey. *LGBTQ Policy Journal at the Harvard Kennedy School*, *2*(1), 13. https://web.archive.org/web/20210719001058/https://escholarship.org/uc/item/2zj46213

H.B. 1346, 123rd Indiana General Assembly (2023). http://web.archive.org/web/20230329111024/https://iga.in.gov/legislative/2023/bills/house/1346

Hedge, C., Powell, G., & Sumner, P. (2017). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

Hekanaho, L. (2020). *Generic and nonbinary pronouns: Usage, acceptability and attitudes* [Doctoral dissertation, University of Helsinki]. HELDA – Digital Repository of the University of Helsinki. http://hdl.handle.net/10138/321581

Henderson, J. M. (2013). Eye movements. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 69–82). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195376746.013.0005

Herman, J. L., Flores, A. R., & O'Neill, K. K. (2022). *How many adults and youth identify as transgender in the United States?* The Williams Institute. https://web.archive.org/web/20230430114921/https://williamsinstitute.law.ucla.edu/publications/trans-adults-united-states/

Hernandez, E. E. (2020). *Pronouns, prescriptivism, and prejudice: Attitudes toward the singular* they*, prescriptive grammar, and nonbinary transgender people* [Master's thesis, Purdue University]. Hammer Research Repository. https://doi.org/10.25394/PGS.12231095.v1

Howe, B. G. (2019). *The impact of microaggressions on transgender identity defense related emotions (TIDE) and the emotional status, desire for social engagement, and participation in suicide-related behaviors in transgender people* (Publication No. 13899488) [Doctoral dissertation, University of Oregon]. ProQuest Dissertations & Theses Global.

Huang, Y. (2006). Coreference: Identity and similarity. In *Encyclopedia of language & linguistics* (pp. 203–205). Elsevier. https://doi.org/10.1016/B0-08-044854-2/00992-5

Human Rights Campaign. (2023, February 15). *Human Rights Campaign working to defeat 340 anti-LGBTQ+ bills at state level already, 150 of which target transgender people – highest number on record*. Human Rights Campaign. https://web.archive.org/web/20230401063456/https://www.hrc.org/press-releases/human-rights-campaign-working-to-defeat-340-anti-lgbtq-bills-at-state-level-already-150-of-which-target-transgender-people-highest-number-on-record

Humane Society. (2020). *Why you should spay/neuter your pet*. http://web.archive.org/web/20201112020818/https://www.humanesociety.org/resources/why-you-should-spayneuter-your-pet

Jacobson, M. B., & Insko, W. R. (1985). Use of nonsexist pronouns as a function of one's feminist orientation. *Sex Roles*, *13*(1–2), 1–7. https://doi.org/10.1007/BF00287456

James, S. E., Herman, J. L., Rankin, S., Keisling, M., Mottet, L. A., & Anafi, M. (2016). *The report of the 2015 U.S. Transgender Survey*. National Center for Transgender Equality. https://web.archive.org/web/20230519030106/https://ustranssurvey.org/reports/#2015report

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*(4), 824–843. https://doi.org/10.1037/0278-7393.20.4.824

Jescheniak, J. D., Schriefers, H., & Lemhöfer, K. (2014). Selection of freestanding and bound gender-marking morphemes in speech production: A review. *Language, Cognition and Neuroscience*, *29*(6), 684–694. https://doi.org/10.1080/01690965.2012.654645

Johnson, A. (2019). Read between the lines: Teaching and learning while nonbinary trans. In G. Beemyn (Ed.), *Trans people in higher education* (pp. 27–34). SUNY Press.

Johnson, I. R., Pietri, E. S., Buck, D. M., & Daas, R. (2021). What's in a pronoun: Exploring gender pronouns as an organizational identity-safety cue among sexual and gender minorities. *Journal of Experimental Social Psychology*, *97*, 104194. https://doi.org/10.1016/j.jesp.2021.104194

Jones, R. P., Cox, D., Fisch-Friedman, M., & Vandermaas-Peeler, A. (2018). *Diversity, division, discrimination: The state of young America*. Public Religion Research Institute. https://web.archive.org/web/20180111030007/https://www.prri.org/research/mtv-culture-and-religion/

Jones, R. P., Jackson, N., Najle, M., Bola, O., & Greenberg, D. (2019). *America's growing support for transgender rights*. Public Religion Research Institute. https://web.archive.org/web/20190611113350/https://www.prri.org/research/americas-growing-support-for-transgender-rights/

Joshi, A. K., Prasad, R., & Miltsakaki, E. (2006). Anaphora resolution: Centering theory approach. In *Encyclopedia of language & linguistics* (pp. 223–230). Elsevier. https://doi.org/10.1016/B0-08-044854-2/04366-2

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*(2), 159–201. https://doi.org/10.1080/016909600386084

Kaiser, E., & Post, C. (2023). *Production of singular* they*: Social biases and gender features on pronouns* [Poster]. 36th Annual Conference on Human Sentence Processing.

Kaiser, E., & Trueswell, J. C. (2011). Investigating the interpretation of pronouns and demonstratives in Finnish: Going beyond salience. In E. Gibson & N. J. Pearlmutter (Eds.), *The processing and acquisition of reference*. MIT Press. https://doi.org/10.7551/mitpress/9780262015127.003.0013

Kattari, S. K., Atteberry-Ash, B., Kinney, M. K., Walls, N. E., & Kattari, L. (2019). One size does not fit all: Differential transgender health experiences. *Social Work in Health Care*, *58*(9), 899–917. https://doi.org/10.1080/00981389.2019.1677279

Kattari, S. K., Kattari, L., Johnson, I., Lacombe-Duncan, A., & Misiolek, B. A. (2020). Differential experiences of mental health among trans/gender diverse adults in Michigan. *International Journal of Environmental Research and Public Health*, *17*(18), 1–16. https://doi.org/doi:10.3390/ijerph17186805

Kattari, S. K., O, A. A., & Kattari, L. (2018). Development and validation of the transgender inclusive behavior scale (TIBS). *Journal of Homosexuality*, *65*(2), 181–196. https://doi.org/10.1080/00918369.2017.1314160

Kehler, A., & Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics*, *39*(1–2), 1–37. https://doi.org/10.1515/tl-2013-0001

Kennison, S. M., Fernandez, E. C., & Bowers, J. M. (2009). Processing differences for anaphoric and cataphoric pronouns: Implications for theories of discourse processing. *Discourse Processes*, *46*(1), 25–45. https://doi.org/10.1080/01638530802359145

Kennison, S. M., & Trofe, J. L. (2003). Comprehending pronouns: A role for word-specific gender stereotype information. *Journal of Psycholinguistic Research*, *23*(3), 355–378. https://doi.org/10.1023/A:1023599719948

Keuroghlian, A. S., Reisner, S. L., White, J. M., & Weiss, R. D. (2015). Substance use and treatment of substance use disorders in a community sample of transgender adults. *Drug and Alcohol Dependence*, *152*, 139–146. https://doi.org/10.1016/j.drugalcdep.2015.04.008

Kidd, K. M., Sequeira, G. M., Douglas, C., Paglisotti, T., Inwards-Breland, D. J., Miller, E., & Coulter, R. W. S. (2021). Prevalence of gender-diverse youth in an urban school district. *Pediatrics*, *147*(6), e2020049823. https://doi.org/10.1542/peds.2020-049823

Koeser, S., Kuhn, E. A., & Sczesny, S. (2015). Just reading? How gender-fair language triggers readers' use of gender-fair forms. *Journal of Language and Social Psychology*, *34*(3), 343–357. https://doi.org/10.1177/0261927X14561119

Koeser, S., & Sczesny, S. (2014). Promoting gender-fair language: The impact of arguments on language use, attitudes, and cognitions. *Journal of Language and Social Psychology*, *33*(5), 548–560. https://doi.org/10.1177/0261927X14541280

Konnelly, L., & Cowper, E. (2020). Gender diversity and morphosyntax: An account of singular *they*. *Glossa: A Journal of General Linguistics*, *5*(1), 1–15. https://doi.org/10.5334/gjgl.1000

Kramer, A., Boland, J., & Queen, R. (2021). *Familiarity is a better predictor than age for specific* they *production, but explicit examples are even better* [Conference talk]. 95th Annual Meeting of the Linguistic Society of America.

Kreiner, H., Sturt, P., & Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language*, *58*(2), 239–261. https://doi.org/10.1016/j.jml.2007.09.003

Kuper, L. E., Nussbaum, R., & Mustanski, B. (2012). Exploring the diversity of gender and sexual orientation identities in an online sample of transgender individuals. *Journal of Sex Research*, *49*(2–3), 244–254. https://doi.org/10.1080/00224499.2011.596954

Lagos, D. (2018). Looking at population health beyond "male" and "female": Implications of transgender identity and gender nonconformity for population health. *Demography*, *55*(6), 2097–2117. https://doi.org/10.1007/s13524-018-0714-3

Lefevor, G. T., Boyd-Rogers, C. C., Sprague, B. M., & Janis, R. A. (2019). Health disparities between genderqueer, transgender, and cisgender individuals: An extension of minority stress theory. *Journal of Counseling Psychology*, *66*(4), 385–395. http://doi.org/10.1037/cou0000339

Levay, K. E., Freese, J., & Druckman, J. N. (2016). The demographic and political composition of Mechanical Turk samples. *SAGE Open*, *6*(1), 1–17. https://doi.org/10.1177/2158244016636433

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–75.

Liceralde, V. R. T., & Watson, D. G. (2023). *Typical psycholinguistic designs require more items to detect reliable individual differences in focal linguistic effects* [Poster]. 36th Annual Conference on Human Sentence Processing.

Lindley, L., & Galupo, M. P. (2020). Gender dysphoria and minority stress: Support for inclusion of gender dysphoria as a proximal stressor. *Psychology of Sexual Orientation and Gender Diversity*, *7*(3), 265–275. http://doi.org/10.1037/sgd0000439

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, *4*(1), 6–14. https://doi.org/10.1016/S1364-6613(99)01418-7

Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998). Eye movements and measures of reading time. In *Eye guidance in reading and scene perception* (pp. 55–75). Elsevier. https://doi.org/10.1016/B978-008043361-5/50004-3

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*, 1171. https://doi.org/10.3389/fpsyg.2015.01171

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

Martin, C. L., Ruble, D. N., & Szkrybalo, J. (2002). Cognitive theories of early gender development. *Psychological Bulletin*, *128*(6), 903–933. https://doi.org/10.1037/0033-2909.128.6.903

Martyna, W. (1980). Beyond the “he/man” approach: The case for nonsexist language. *Signs: Journal of Women in Culture and Society*, *5*(3), 482–493. https://doi.org/10.1086/493733

McConnell-Ginet, S. (2014). Gender and its relation to sex: The myth of “natural” gender. In G. G. Corbett (Ed.), *The expression of gender* (pp. 3–38). https://doi.org/10.1515/9783110307337.3

McGonagill, R. (2023). *Job-seekers with nonbinary gender pronouns on their resumes are less likely to be contacted by employers*. Business.com. https://web.archive.org/web/20230308193858/https:/www.business.com/hiring/nonbinary-discrimination-job-market-report/

McKoon, G., Gerrig, R. J., & Greene, S. B. (1996). Pronoun resolution without pronouns: Some consequences of memory-based text processing. *Journal of Experimental Psychology: Learning Memory and Cognition*, *22*(4), 919–932. https://doi.org/10.1037/0278-7393.22.4.919

McLemore, K. A. (2015). Self and identity experiences with misgendering: Identity misclassification of transgender spectrum individuals. *Self and Identity*, *14*(1), 51–74. https://doi.org/10.1080/15298868.2014.950691

McLemore, K. A. (2018). A minority stress perspective on transgender individuals’ experiences with misgendering. *Stigma and Health*, *3*(1), 53–64. https://doi.org/10.1037/sah0000070

McMinn, M. R., Lindsay, S. F., Hannum, L. E., & Troyer, P. K. (1990). Does sexist language reflect personal characteristics? *Sex Roles*, *23*(7–8), 389–396. https://doi.org/10.1007/BF00289227

McMurray, B., Klein-Packard, J., & Tomblin, J. B. (2019). A real-time mechanism underlying lexical deficits in developmental language disorder: Between-word inhibition. *Cognition*, *191*, 104000. https://doi.org/10.1016/j.cognition.2019.06.012

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category VOT affects recovery from “lexical” garden-paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language*, *60*(1), 65–91. https://doi.org/10.1016/j.jml.2008.07.002

Meerwijk, E. L., & Sevelius, J. M. (2017). Transgender population size in the United States: A meta-regression of population-based probability samples. *AJPH Transgender Health*, *107*(2), 1–8. https://doi.org/10.2105/AJPH.2016.303578

Minkin, R., & Brown, A. (2021, July 27). *Rising shares of U.S. adults know someone who is transgender or goes by gender-neutral pronouns*. Pew Research Center. https://web.archive.org/web/20210731145136/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/

Mitchell, L., MacArthur, H. J., & Blomquist, K. K. (2021). The effect of misgendering on body dissatisfaction and dietary restraint in transgender individuals: Testing a misgendering-congruence process. *International Journal of Eating Disorders*, *54*(7), 1295–1301. https://doi.org/10.1002/eat.23537

Molinaro, N., Su, J. J., & Carreiras, M. (2016). Stereotypes override grammar: Social knowledge in sentence comprehension. *Brain and Language*, *155–156*, 36–43. https://doi.org/10.1016/j.bandl.2016.03.002

Monáe, J., & Lore, D. (2022). Nevermind. In *The memory librarian: And other stories of dirty computer*. Harper Voyager.

Morehouse, K. N., Kurdi, B., Hakim, E., & Banaji, M. R. (2022). When a stereotype dumbfounds: Probing the nature of the surgeon = male belief. *Current Research in Ecological and Social Psychology*, *3*, 100044. https://doi.org/10.1016/j.cresp.2022.100044

Morris, E. R., & Galupo, M. P. (2019). "Attempting to dull the dysphoria": Nonsuicidal self-injury among transgender individuals. *Psychology of Sexual Orientation and Gender Diversity*, *6*(3), 296–307. http://doi.org/10.1037/sgd0000327

Nadal, K. L. (2013). Gender identity microaggressions: Experiences of transgender and gender nonconforming people. In *That's so gay! Microaggressions and the lesbian, gay, bisexual, and transgender community* (pp. 80–107). American Psychological Association. https://doi.org/10.1037/14093-005

Nadal, K. L. (2019). Measuring LGBTQ microaggressions: The sexual orientation microaggressions scale (SOMS) and the gender identity microaggressions Scale (GIMS). *Journal of Homosexuality*, *66*(10), 1404–1414. https://doi.org/10.1080/00918369.2018.1542206

Nadal, K. L., Davidoff, K. C., Davis, L. S., & Wong, Y. (2014). Emotional, behavioral, and cognitive reactions to microaggressions: Transgender perspectives. *Psychology of Sexual Orientation and Gender Diversity*, *1*(1), 72–81. https://doi.org/10.1037/sgd0000011

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

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

National Academies of Sciences, Engineering, and Medicine. (2022). *Measuring sex, gender identity, and sexual orientation*. National Academies Press. https://doi.org/10.17226/26424

Neenan, M. (2020, February 3). *Other Voices: Matthew Neenan* [Video]. Nashville Ballet. https://www.youtube.com/watch?v=6BhPDXxB4Tg

Newcomb, M. E., Hill, R., Buehler, K., Ryan, D. T., Whitton, S. W., & Mustanski, B. (2020). High burden of mental health problems, substance use, violence, and related psychosocial factors in transgender, non-binary, and gender diverse youth and young adults. *Archives of Sexual Behavior*, *49*(2), 645–659. https://doi.org/10.1007/s10508-019-01533-9

Nichols, A., Chun, E., & Almor, A. (2019). *Non-binary gendered reference in LGBTQ+ English: Implications of singular* they *for sentence processing* [Poster]. 32nd Annual Conference on Human Sentence Processing.

Nieuwland, M. S. (2014). "Who's he?" Event-related brain potentials and unbound pronouns. *Journal of Memory and Language*, *76*, 1–28. https://doi.org/10.1016/j.jml.2014.06.002

Noonan, P. (2019, July 25). What were Robespierre's pronouns? *The Wall Street Journal*. https://web.archive.org/web/20210221063231/https://www.wsj.com/articles/what-were-robespierres-pronouns-11564095088

Oakhill, J., Garnham, A., & Reynolds, D. (2005). Immediate activation of stereotypical gender information. *Memory & Cognition*, *33*(6), 972–983. https://doi.org/10.3758/BF03193206

OED Online. (2021). They*, pron., adj., adv.,* and *n.* In *Oxford English dictionary*. Oxford University Press. https://web.archive.org/web/20210628005850/https:/www.oed.com/view/Entry/200700

Osterhout, L., Bersick, M., & Mclaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition*, *25*(3), 273–285. https://doi.org/10.3758/BF03211283

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785–806. https://doi.org/10.1016/0749-596X(92)90039-Z

Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*(4), 786–803. https://doi.org/10.1037/0278-7393.20.4.786

Parker, K., Graf, N., & Igielnik, R. (2019). *Generation Z looks a lot like millennials on key social and political issues*. Pew Research Center. https://web.archive.org/web/20210210083104/https:/www.pewresearch.org/social-trends/2019/01/17/generation-z-looks-a-lot-like-millennials-on-key-social-and-political-issues/

Parks, C., Guskin, E., & Clement, S. (2023, March 23). Most trans adults say transitioning made them more satisfied with their lives. *The Washington Post*. https://web.archive.org/web/20230323123428/https:/www.washingtonpost.com/dc-md-va/2023/03/23/transgender-adults-transitioning-poll/

Parks, J. B., & Roberton, M. A. (1998). Contemporary arguments against nonsexist language: Blaubergs (1980) revisited. *Sex Roles*, *39*(5–6), 445–461. https://doi.org/10.1023/A:1018827227128

Parks, J. B., & Roberton, M. A. (2004). Attitudes toward women mediate the gender effect on attitudes toward sexist language. *Psychology of Women Quarterly*, *28*(3), 233–239. https://doi.org/10.1111/j.1471-6402.2004.00140.x

Parks, J. B., & Roberton, M. A. (2008). Generation gaps in attitudes toward sexist/nonsexist language. *Journal of Language and Social Psychology*, *27*(3), 276–283. https://doi.org/10.1177/0261927X08317956

Peer, E., Rothschild, D., Gordon, G., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, *54*(4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

Penelope, J. (1982). Power and the opposition to feminist proposals for language change. *College English*, *44*(8), 840–848. https://doi.org/10.2307/377340

Penklis, H. S. (2020). *"Celebrated, not tolerated": How transgender and nonbinary individuals navigate the social world* (Publication No. 27999909) [Master's thesis, The University of Texas at El Paso]. ProQuest Dissertations & Theses Global.

Perez-Brumer, A., Day, J. K., Russell, S. T., & Hatzenbuehler, M. L. (2017). Prevalence and correlates of suicidal ideation among transgender youth in California: Findings from a representative, population-based sample of high school students. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*(9), 739–746. https://doi.org/10.1016/j.jaac.2017.06.010

Peterson, J. (2016, September 30). “I’m not a bigot” Meet the U of T prof who refuses to use genderless pronouns. *CBC As It Happens*. https://web.archive.org/web/20210506213642/https://www.cbc.ca/radio/asithappens/as-it-happens-friday-edition-1.3786140/i-m-not-a-bigot-meet-the-u-of-t-prof-who-refuses-to-use-genderless-pronouns-1.3786144%0A

Pitcher, E. N. (2017). “There’s stuff that comes with being an unexpected guest”: Experiences of trans* academics with microaggressions. *International Journal of Qualitative Studies in Education*, *30*(7), 688–703. https://doi.org/10.1080/09518398.2017.1309588

Pozzan, L., & Antón-Méndez, I. (2017). English possessive gender agreement in production and comprehension: Similarities and differences between young monolingual English learners and adult Mandarin-English second language learners. *Applied Psycholinguistics*, *38*(4), 985–1017. https://doi.org/10.1017/S0142716417000017

Prasad, G., & Morris, J. (2020). *The P600 for singular* they*: How the brain reacts when John decides to treat themselves to sushi* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/hwzke

Protection Against Child Medical and Psychological Experimentation Act, H.B. 619, 168th New Hampshire General Court, (2023). http://web.archive.org/web/20230322225529/https:/gencourt.state.nh.us/bill_Status/billinfo.aspx?id=71

Public Religion Research Institute. (2021). *Most Americans support pro-LGBTQ policies, but are more divided over transgender sports and bathroom policies.* https://web.archive.org/web/20211001032111/https:/www.prri.org/research/most-americans-support-pro-lgbtq-policies-but-are-more-divided-over-transgender-sports-and-bathroom-policies/

Puckett, J. A., Maroney, M. R., Wadsworth, L. P., Mustanski, B., & Newcomb, M. E. (2020). Coping with discrimination: The insidious effects of gender minority stigma on depression and anxiety in transgender individuals. *Journal of Clinical Psychology*, *76*(1), 176–194. https://doi.org/10.1002/jclp.22865

Puckett, J. A., Matsuno, E., Dyar, C., Mustanski, B., & Newcomb, M. E. (2019). Mental health and resilience in transgender individuals: What type of support makes a difference? *Journal of Family Psychology*, *33*(8), 954–964. https://doi.org/10.1037/fam0000561

Pyykkönen, P., Hyönä, J., & Van Gompel, R. P. G. (2010). Activating gender stereotypes during online spoken language processing: Evidence from visual world eye tracking. *Experimental Psychology*, *57*(2), 126–133. https://doi.org/10.1027/1618-3169/a000016

R Core Team. (2023). *R: A language and environment for statistical computing* (Version 4.2.3). R Foundation for Statistical Computing. https://www.r-project.org/

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). *Robust speech recognition via large-scale weak supervision* [Preprint]. arXiv. https://doi.org/10.48550/ARXIV.2212.04356

Rayner, K. (1988). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Rayner, K., & Juhasz, B. (2006). Reading processes in adults. In *Encyclopedia of language & linguistics* (pp. 373–378). Elsevier. https://doi.org/10.1016/B0-08-044854-2/00794-X

Reali, C., Esaulova, Y., & Von Stockhausen, L. (2015). Isolating stereotypical gender in a grammatical gender language: Evidence from eye movements. *Applied Psycholinguistics*, *36*(4), 977–1006. https://doi.org/10.1017/S0142716414000010

Reynolds, D. J., Garnham, A., Oakhill, J., & Reynolds, D. J. (2006). Evidence of immediate activation of gender information from a social role name. *The Quarterly Journal of Experimental Psychology*, *59*(5), 886–903. https://doi.org/10.1080/02724980543000088

Richards, C., & Barker, M. (2013). Further genders. In *Sexuality & gender for mental health professionals: A practical guide*. SAGE Publications Ltd. https://doi.org/10.4135/9781473957817

Robertson, M. (2021). *Breaking, bending, and stretching the rules of singular* they [Conference talk]. 27th Annual Lavender Languages and Linguistics Conference.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*, 107–142. https://doi.org/10.1016/0010-0277(92)90041-F

Rood, B. A., Reisner, S. L., Surace, F. I., Puckett, J. A., Maroney, M. R., & Pantalone, D. W. (2016). Expecting rejection: Understanding the minority stress experiences of transgender and gender-nonconforming individuals. *Transgender Health*, *1*(1), 151–164. https://doi.org/10.1089/trgh.2016.0012

Rubin, D. L., Greene, K., & Schneider, D. (1994). Adopting gender-inclusive language reforms: Diachronic and synchronic variation. *Journal of Language and Social Psychology*, *13*(2), 91–114. https://doi.org/10.1177/0261927X94132001

Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2006). Processing reflexives and pronouns in picture noun phrases. *Cognitive Science*, *30*(2), 193–241. https://doi.org/10.1207/s15516709cog0000_58

Saltzburg, S., & Davis, T. S. (2010). Co-authoring gender-queer youth identities: Discursive *tellings* and *retellings*. *Journal of Ethnic And Cultural Diversity in Social Work*, *19*(2), 87–108. https://doi.org/10.1080/15313200903124028

Samar, V. J. (2006). Evoked potentials. In *Encyclopedia of language & linguistics* (pp. 326–335). Elsevier. https://doi.org/10.1016/B0-08-044854-2/01089-0

Sanford, A. J. (1985). *Cognition and cognitive psychology*. Weidenfeld and Nicolson.

Sanford, A. J., & Filik, R. (2007). *They* as a gender-unspecified singular pronoun: Eye tracking reveals a processing cost. *Quarterly Journal of Experimental Psychology*, *60*(2), 171–178. https://doi.org/10.1080/17470210600973390

Sanford, A. J., & Garrod, S. C. (1981). *Understanding written language: Explorations of comprehension beyond the sentence*. Wiley.

Sarrasin, O., Gabriel, U., & Gygax, P. (2012). Sexism and attitudes toward gender-neutral language: The case of English, French, and German. *Swiss Journal of Psychology*, *71*(3), 113–124. http://doi.org/10.1024/1421-0185/a000078

Sato, S., Gygax, P., & Gabriel, U. (2013). Gender inferences: Grammatical features and their impact on the representation of gender in bilinguals. *Bilingualism: Language and Cognition*, *16*(4), 792–807. https://doi.org/10.1017/S1366728912000739

S.B. 30, 59th Oklahoma Legislature (2022). http://web.archive.org/web/20230331001715/http:/www.oklegislature.gov/BillInfo.aspx?Bill=sb30&Session=2300

S.B. 354, 123rd Indiana General Assembly (2023). http://web.archive.org/web/20230311015856/https:/iga.in.gov/legislative/2023/bills/senate/354

S.B. 0466, H.B. 1269, 113th Tennessee General Assembly (2023). http://web.archive.org/web/20230317012106/https:/wapp.capitol.tn.gov/apps/BillInfo/Default.aspx?BillNumber=SB0466

S.B. 1001, 56th Arizona State Legislature (2022). http://web.archive.org/web/20230313195640/https:/www.azleg.gov/legtext/56leg/1R/bills/SB1001P.htm

S.B. 1700, 56th Arizona State Legislature (2023). http://web.archive.org/web/20230313195640/https:/www.azleg.gov/legtext/56leg/1R/bills/SB1700P.htm

S.B. 2231, 68th North Dakota Legislative Assembly (2023). http://web.archive.org/web/20230319235209/https:/www.ndlegis.gov/assembly/68-2023/regular/bill-index/bi2231.html

S.B. 2260, 68th North Dakota Legislative Assembly (2023). http://web.archive.org/web/20230129022109/https:/www.ndlegis.gov/assembly/68-2023/regular/bill-index/bi2260.html

Schiller, N. O., & Caramazza, A. (2003). Grammatical feature selection in noun phrase production: Evidence from German and Dutch. *Journal of Memory and Language*, *48*(1), 169–194. https://doi.org/10.1016/S0749-596X(02)00508-9

Schmitt, B. M., Lamers, M., & Munte, T. F. (2002). Electrophysiological estimates of biological and syntactic gender violation during pronoun processing. *Cognitive Brain Research*, *14*(3), 333–346. https://doi.org/10.1016/S0926-6410(02)00136-2

Schmitt, B. M., Meyer, A. S., & Levelt, W. J. M. (1999). Lexical access in the production of pronouns. *Cognition*, *69*(3), 313–335. https://doi.org/10.1016/S0010-0277(98)00073-0

School Establishment Clause Act, S.B. 937, 59th Oklahoma Legislature (2023). http://web.archive.org/web/20230129014816/http://www.oklegislature.gov/BillInfo.aspx?Bill=SB937

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through context representation. *Cognition*, *71*(2), 109–147. https://doi.org/10.1016/S0010-0277(99)00025-6

Shakespeare, W. (1623). *A comedy of errors*.

Shenkar, E., Mansoor, N., Ahn, J. S., Owen-Oliner, L., Chen, P., & Grodner, D. (2023). *Investigating the processing of singular* they *using a web-based maze task* [Conference talk]. 36th Annual Conference on Human Sentence Processing.

Shields, J. P., Cohen, R., Glassman, J. R., Whitaker, K., Franks, H., & Bertolini, I. (2013). Estimating population size and demographic characteristics of lesbian, gay, bisexual, and transgender youth in middle school. *Journal of Adolescent Health*, *52*(2), 248–250. https://doi.org/10.1016/j.jadohealth.2012.06.016

Siewierska, A. (2013). Gender distinctions in independent personal pronouns. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology. https://web.archive.org/web/20230309235830/https:/wals.info/chapter/44

Sims, J. (Author), & Newall, A. J. (Producer/Director). (2019, June 27). *Scrutiny (No. 142)* [Audio podcast episode]. In *The Magnus archives.* Rusty Quill. https://play.acast.com/s/themagnusarchives/mag142-scrutiny

Sims, J. (Author), & Newall, A. J. (Producer/Director). (2020, June 18). *Strung out (No. 172)* [Audio podcast episode]. In *The Magnus archives.* Rusty Quill. https://play.acast.com/s/themagnusarchives/mag172-strungout

Slobin, D. I. (1996). From “thought and language” to “thinking for speaking.” In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge University Press.

Snow, B. (Author), & Shippen, L. (Producer/Director). (2021, August 11). *Infernum septimana (No. 7.05)* [Audio podcast episode]. In *The college tapes.* Atypical Artists. https://www.thebrightsessions.com/episode-705

Song, H., & Fisher, C. (2005). Who's "she"? Discourse prominence influences preschoolers' comprehension of pronouns. *Journal of Memory and Language*, *52*(1), 29–57. https://doi.org/10.1016/j.jml.2004.06.012

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, *15*(1), 72–101. https://doi.org/10.2307/1412159

Speyer, L. G., & Schleef, E. (2019). Processing "gender-neutral" pronouns: A self-paced reading study of learners of English. *Applied Linguistics*, *40*(5), 793–815. https://doi.org/10.1093/applin/amy022

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, *45*(4), 447–481. https://doi.org/10.1016/S0010-0285(02)00503-0

Staub, A. (2021). How reliable are individual differences in eye movements in reading? *Journal of Memory and Language*, *116*, 104190. https://doi.org/10.1016/j.jml.2020.104190

Stauffer, R. (2021, June). The one-size-fits-all narrative of your 20s needs to change. *The Atlantic*. https://web.archive.org/web/20210621153724/https:/www.theatlantic.com/ideas/archive/2021/06/your-20s-dont-have-to-be-the-best-time-of-your-life/619187/

Sterzing, P. R., Ratliff, G. A., Gartner, R. E., McGeough, B. L., & Johnson, K. C. (2017). Social ecological correlates of polyvictimization among a national sample of transgender, genderqueer, and cisgender sexual minority adolescents. *Child Abuse & Neglect*, *67*, 1–12. https://doi.org/10.1016/j.chiabu.2017.02.017

Streed Jr., C. G., McCarthy, E. P., & Haas, J. S. (2017). Association between gender minority status and self-reported physical and mental health in the United States. *JAMA Internal Medicine*, *177*(8), 2015–2017. https://doi.org/10.1001/jamainternmed.2017.1460

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, *48*(3), 542–562. https://doi.org/10.1016/S0749-596X(02)00536-3

Sue, D. W., Capodilupo, C. M., Torino, G. C., Bucceri, J. M., Holder, A. M. B. B., Nadal, K. L., & Esquilin, M. (2007). Racial microaggressions in everyday life: Implications for clinical practice. *American Psychologist*, *62*(4), 271–286. https://doi.org/10.1037/0003-066X.62.4.271

Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. (1995). Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology*, *68*(2), 199–214. https://doi.org/10.1037//0022-3514.68.2.199

Swim, J. K., Mallett, R., & Stangor, C. (2004). Understanding subtle sexism: Detection and use of sexist language. *Sex Roles*, *51*(3/4), 117–128. https://doi.org/10.1023/B:SERS.0000037757.73192.06

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. G. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, *29*(6), 557–580. https://doi.org/10.1023/A:1026464108329

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634. https://doi.org/10.1126/science.7777863

Tebbe, E. A., & Moradi, B. (2016). Suicide risk in trans populations: An application of minority stress theory. *Journal of Counseling Psychology*, *63*(5), 520–533. http://doi.org/10.1037/cou0000152

Tebbe, E. N., & Moradi, B. (2012). Anti-transgender prejudice: A structural equation model of associated constructs. *Journal of Counseling Psychology*, *59*(2), 251–261. http://doi.org/10.1037/a0026990

Testa, R. J., Habarth, J., Peta, J., Balsam, K., & Bockting, W. (2015). Development of the gender minority stress and resilience measure. *Psychology of Sexual Orientation and Gender Diversity*, *2*(1), 65–77. https://doi.org/10.1037/sgd0000081

The Trevor Project. (2020). *2020 National survey on LGBTQ youth mental health*. https://web.archive.org/web/20210104100949/https://www.thetrevorproject.org/survey-2020/

The Trevor Project. (2022). *2022 National survey on LGBTQ youth mental health*. https://web.archive.org/web/20230505165356/https://www.thetrevorproject.org/survey-2022/

Towers, A. (2021, July 2). Demi Lovato thanks Lizzo for correcting paparazzo who misgendered her: “I love you.” *Entertainment Weekly*. https://web.archive.org/web/20210702144731/https:/ew.com/music/demi-lovato-thanks-lizzo-correcting-paparazzo-who-misgendered-her/

Truszczynski, N., Singh, A. A., & Hansen, N. (2020). The discrimination experiences and coping responses of non-binary and trans people. *Journal of Homosexuality*, *69*(4), 741–755. https://doi.org/10.1080/00918369.2020.1855028

Twenge, J. M. (2023, May 1). How Gen Z changed its views on gender. *Time*. https://web.archive.org/web/20230501110039/https://time.com/6275663/generation-z-gender-identity/

UK Government Equalities Office. (2018). *National LGBT survey*. https://web.archive.org/web/20210915151037/https:/www.gov.uk/government/publications/national-lgbt-survey-summary-report

Valentine, S. E., & Shipherd, J. C. (2018). A systematic review of social stress and mental health among transgender and gender non-conforming people in the United States. *Clinical Psychology Review*, *66*(2017), 24–38. https://doi.org/10.1016/j.cpr.2018.03.003

Veale, J. F., Peter, T., Travers, R., & Saewyc, E. M. (2017). Enacted stigma, mental health, and protective factors among transgender youth in Canada. *Transgender Health*, *2*(1), 207–216. https://doi.org/10.1089/trgh.2017.0031

Veale, J. F., Watson, R. J., Peter, T., & Saewyc, E. M. (2017). Mental health disparities among Canadian transgender youth. *Journal of Adolescent Health*, *60*(1), 44–49. https://doi.org/10.1016/ j.jadohealth.2016.09.014

Voeten, C. C. (2023). *buildmer: Stepwise elimination and term reordering for mixed-effects regression* (Version 2.8) [R]. https://CRAN.R-project.org/package=buildmer

von Restorff, H. (1933). Uber die wirkung von bereichsbildungen im spurenfeld [On the effect of field formations in the trace field]. *Psychologische Forschung*, *18*, 299–342.

Walker, M. A. (2002). Toward a model of the interaction of centering with global discourse structure. *Verbum*, *1*, 31–58.

Wallace, W. P. (1965). Review of the historical, empirical, and theoretical status of the von Restorff phenomenon. *Psychological Bulletin*, *63*(6), 410–424. https://doi.org/10.1037/h0022001

Wang, M., & Schiller, N. O. (2019). A review on grammatical gender agreement in speech production. *Frontiers in Psychology*, *9*, 2754. https://doi.org/10.3389/fpsyg.2018.02754

Waxman, S. R. (2010). Names will never hurt me? Naming and the development of racial and gender categories in preschool-aged children. *European Journal of Social Psychology*, *40*(4), 593–610. https://doi.org/10.1002/ejsp.732

Weinhardt, L. S., Xie, H., Wesp, L. M., Murray, J. R., Apchemengish, I., Kioko, D., Weinhardt, C. B., & Cook-Daniels, L. (2019). The role of family, friend, and significant other support in well-being among transgender and non-binary youth. *Journal of GLBT Family Studies*, *15*(4), 311–325. https://doi.org/10.1080/1550428X.2018.1522606

White Hughto, J. M., Reisner, S. L., & Pachankis, J. E. (2015). Transgender stigma and health: A critical review of stigma determinants, mechanisms, and interventions. *Social Science & Medicine*, *145*, 222–231. https://doi.org/10.1016/j.socscimed.2015.11.010

White, R. G. (1880). *Everyday English*. Houghton, Mifflin and Company.

Wilson, B. D. M., & Meyer, I. H. (2021). *Nonbinary LGBTQ adults in the United States*. The Williams Institute. https://web.archive.org/web/20210623183643/https:/williamsinstitute.law.ucla.edu/publications/nonbinary-lgbtq-adults-us/

Zehr, J., & Schwartz, F. (2018). *PennController for internet-based experiments (IBEX)*. https://doi.org/10.17605/OSF.IO/MD832

Zimman, L. (2017). Transgender language reform: Some challenges and strategies for promoting trans-affirming, gender-inclusive language. *Journal of Language and Discrimination*, *1*(1), 84–105. https://doi.org/10.1558/jld.33139

APPENDICES

LIST OF SUPPLEMENTARY TABLES

LIST OF SUPPLEMENTARY FIGURES

APPENDIX A

Supplementary Analyses

A.1 Experiment 1

A.1.1 Experiment 1A: Additional Results

In addition to whether the characters used he/him, she/her, or they/them, participants were also asked about the characters' jobs and pets (Figure A1.1). Accuracy matching the 12 jobs to the 12 characters was lower (M = 0.21, SD = 0.41) than for the 3 pets (M = 0.41, SD = 0.49) and 3 pronouns (M = 0.66, SD = 0.47), but not at floor. Neither job nor pet accuracy varied based on the character's pronouns. Accuracy for the characters' pets (cat, dog, or fish) was designed as a comparison to pronoun accuracy, and the two were compared in a model including the Character's Pronouns (contrast coded as in the main analyses) and Question Type (pronoun vs pet, mean-center effects coded) as fixed effects. The most complex model that converged included by-participant random slopes for Question Type (Table A1.1). Averaging across the three character pronouns, participants were significantly more accurate for pronoun questions than pet questions (β = 1.17, z = 11.19, $p < .001$). The interaction between Character Pronoun (they/them vs he/him + she/her) and Question Type was significant (β = 1.45, z = 7.55, $p < .001$), reflecting that the character pronouns affected accuracy for the pronoun question, but not the pet question. Probing this interaction indicated that for they/them characters, there was no significant difference in accuracy between pronouns and pets (β = 0.21, z = 1.35, $p = .18$), but that for he/him + she/her characters, pronoun accuracy was higher than pet accuracy (β = 1.65, z = 12.97, $p < .001$).

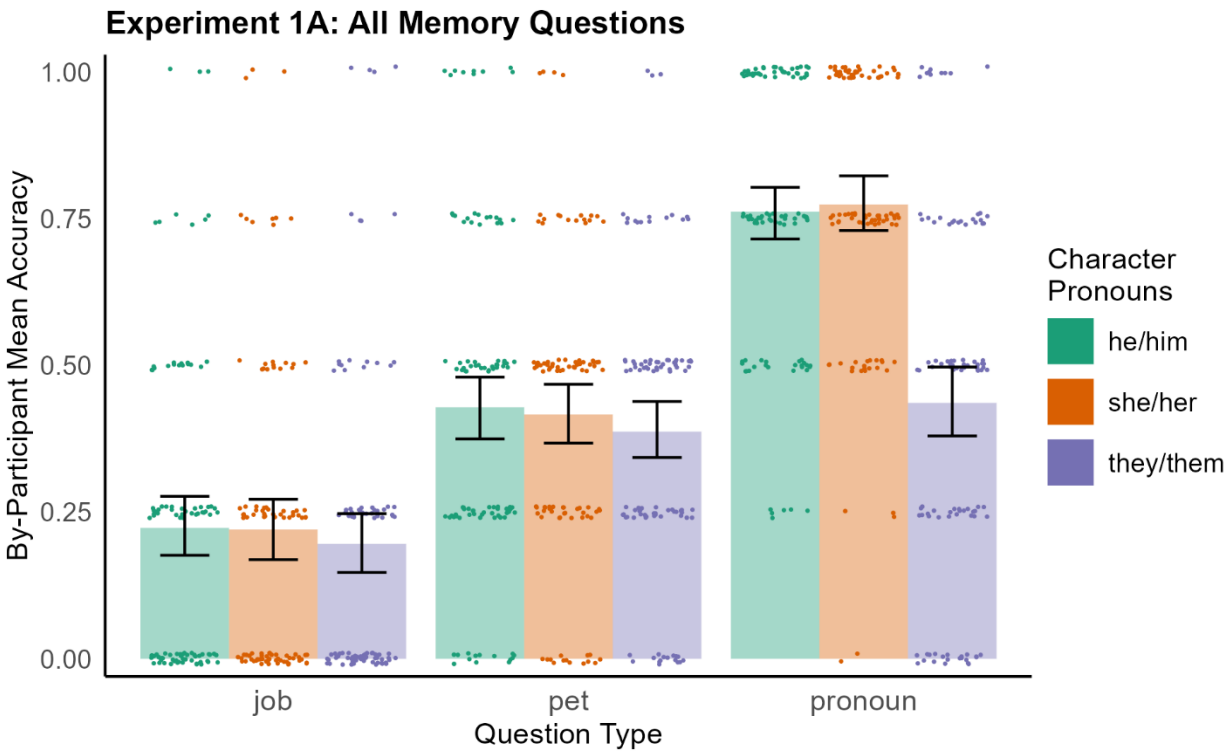


Figure A1.1: Experiment 1A: All Memory Questions. By-participant mean accuracy in the multiple-choice memory task for each character's pronouns, pet, and job, with colors indicating the character's pronouns. Error bars indicate 95% CIs calculated over the by-participant means.

Table A1.1: Experiment 1A: Comparing Pronoun Accuracy to Pet Accuracy. Model results for the effects of Character Pronoun and Question Type (character's pronoun or pet) on Memory Accuracy. Character Pronoun is contrast-coded as in the main analysis.

Experiment 1A: Memory for Pronouns vs Pets

	Memory Accuracy			
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	0.199	0.073	2.731	**0.006**
Question Type: Pet (-.5) **vs Pronoun** (+.5)	1.168	0.104	11.189	**<0.001**
Character Pronoun: They (-.66) **vs He** (+.33) **+ She** (+.33)	0.884	0.096	9.216	**<0.001**
Character Pronoun: He (-.5) vs She (+.5)	0.010	0.113	0.088	0.930
Question Type * Character Pronoun (They vs He + She)	1.449	0.192	7.548	**<0.001**
Question Type * Character Pronoun (He vs She)	0.128	0.226	0.566	0.571
Random Effects				
$T_{00\ Participant}$	0.323			
$T_{11\ Question\ Type\ |\ Participant}$	0.243			
$\rho_{01\ Participant}$	0.231			
$N_{Participant}$	102			
Observations	2448			

The memory and production tasks were compared directly by creating a model predicting accuracy in both tasks, with Task as a mean-center effects coded fixed effect (Table A1.2). The main effect of Task was not significant (β = 0.11, z = 1.04, p = .30), but the interaction between Pronoun (They vs He + She) was significant (β = 1.38, z = 6.36, p < .001). Probing this interaction indicated that memory was more accurate than production for they/them characters (β = -0.80, z = -4.77, p < .001). Conversely, memory was less accurate than production for he/him and she/her characters (β = 0.56, z = 4.19, p < .001).

Table A1.2: Experiment 1A: Comparing Memory & Production. Model results for the effects of Pronoun and Task (memory vs production) on Accuracy.

Experiment 1A: Comparing Memory and Production Accuracy

	Accuracy			
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	0.891	0.092	9.723	**<0.001**
Pronoun: They (-.66) **vs He** (+.33) **+ She** (+.33)	2.469	0.204	12.104	**<0.001**
Pronoun: He (-.5) vs She (+.5)	0.155	0.172	1.041	0.368
Task: Memory (-.5) vs Production (+.5)	0.110	0.105	1.041	0.298
Pronoun (They vs He + She) * Task	1.380	0.217	6.357	**<0.001**
Pronoun (He vs She) * Task	0.152	0.268	0.568	0.570
Random Effects				
$\tau_{00\ Participant}$	0.468			
$\tau_{11\ Pronoun\ (They\ vs\ He\ +\ She)\ |\ Participant}$	2.737			
$\tau_{11\ Pronoun\ (He\ vs\ She)\ |\ Participant}$	0.356			
$\rho_{01\ Pronoun\ (They\ vs\ He\ +\ She)\ |\ Participant}$	0.091			
$\rho_{01\ Pronoun\ (He\ vs\ She)\ |\ Participant}$	0.098			
$N_{\ Participant}$	102			
Observations	2448			

A.1.2 Experiment 1B: Additional Results

Table A1.3: Experiment 1B: Memory. Model results for the effect of Pronoun on Memory Accuracy, when the memory task was completed after the production task.

Experiment 1B: Memory

	Memory Accuracy			
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	0.949	0.095	10.028	**<0.001**
Pronoun: They (-.66) **vs He** (+.33) **+ She** (+.33)	1.552	0.143	10.882	**<0.001**
Pronoun: He (-.5) vs She (+.5)	0.096	0.179	0.540	0.589
Random Effects				
$T_{00\ Participant}$	0.380			
$N_{Participant}$	101			
Observations	1212			

Table A1.4: Experiment 1B: Production. Model results for the effect of Pronoun on Production Accuracy, when the memory task was completed after the production task.

Experiment 1B: Production

	Production Accuracy			
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	1.108	0.123	8.974	**<0.001**
Pronoun: They (-.66) **vs He** (+.33) **+ She** (+.33)	2.476	0.167	14.788	**<0.001**
Pronoun: He (-.5) vs She (+.5)	0.002	0.224	0.009	0.993
Random Effects				
$T_{00\ Participant}$	0.779			
$T_{00\ Name}$	0.008			
$N_{Participant}$	101			
N_{Name}	12			
Observations	1212			

Table A1.5: Experiment 1B: Memory Predicting Production. Model results for the effects of Pronoun and Memory Accuracy on Production Accuracy, when the memory task was completed after the production task.

Experiment 1B: Memory Predicting Production

	Production Accuracy			
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	0.767	0.080	9.538	**<0.001**
Pronoun: They (-.66) **vs He** (+.33) **+ She** (+.33)	1.991	0.161	12.349	**<0.001**
Pronoun: He (-.5) vs She (+.5)	-0.040	0.209	-0.194	0.846
Memory Accuracy: Wrong (-.5) **vs Right** (+.5)	1.491	0.161	9.274	**<0.001**
Pronoun (They vs He + She) * Memory Accuracy	-1.156	0.322	-3.584	**<0.001**
Pronoun (He vs She) * Memory Accuracy	0.240	0.418	0.575	0.565
Observations	1212			

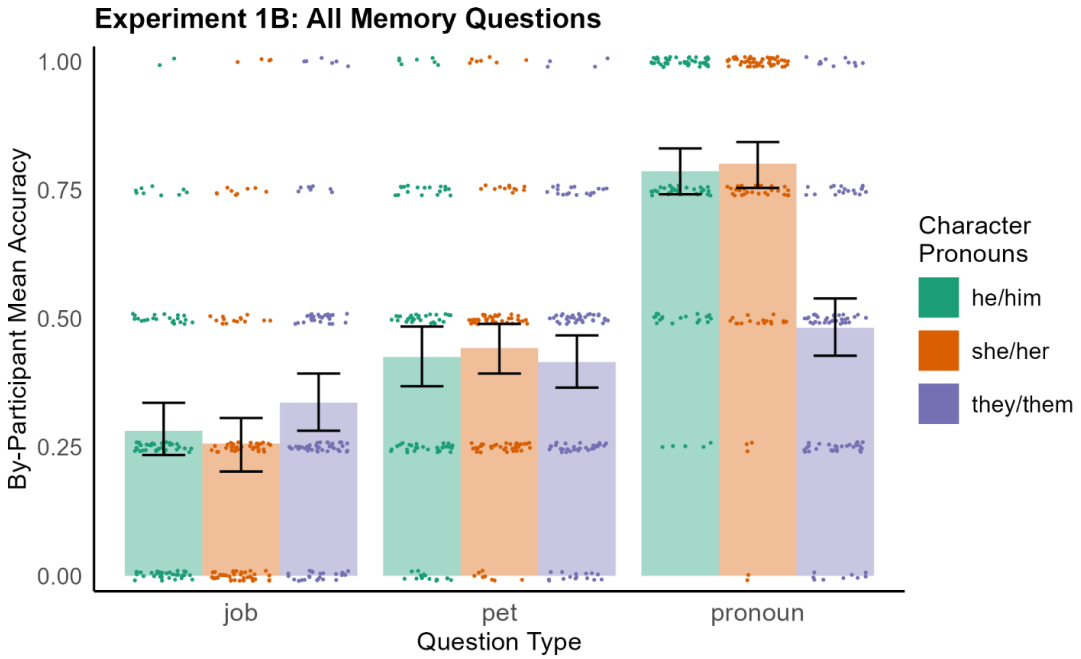


Figure A1.2: Experiment 1B: All Memory Questions. By-participant mean accuracy in the multiple-choice memory task for each character's pronouns, pet, and job, with colors indicating the character's pronouns. Error bars indicate 95% CIs calculated over the by-participant means.

Table A1.6: Experiment 1B: Comparing Pronoun Accuracy to Pet Accuracy. Model results for the effects of Character Pronoun and Question Type (character's pronoun or pet) on Memory Accuracy, when the memory task was completed after the production task.

Experiment 1B: Memory for Pronouns vs Pets

	Memory Accuracy			
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	0.324	0.072	4.505	**<0.001**
Question Type: Pet (-.5) **vs Pronoun** (+.5)	1.255	0.103	12.217	**<0.001**
Character Pronoun: They (-.66) **vs He** (+.33) **+ She** (+.33)	0.818	0.096	8.532	**<0.001**
Character Pronoun: He (-.5) vs She (+.5)	0.075	0.124	0.600	0.549
Question Type * Character Pronoun (They vs He + She)	1.474	0.192	7.688	**<0.001**
Question Type * Character Pronoun (He vs She)	0.022	0.231	0.095	0.924
Random Effects				
$\tau_{00\ Participant}$	0.256			
$\tau_{00\ Name}$	0.005			
$\tau_{11\ Question\ Type\ \mid\ Participant}$	0.177			
$\rho_{01\ Participant}$	0.416			
$N_{Participant}$	101			
N_{Name}	12			
Observations	2424			

A.1.3 Comparing Experiments 1A & 1B

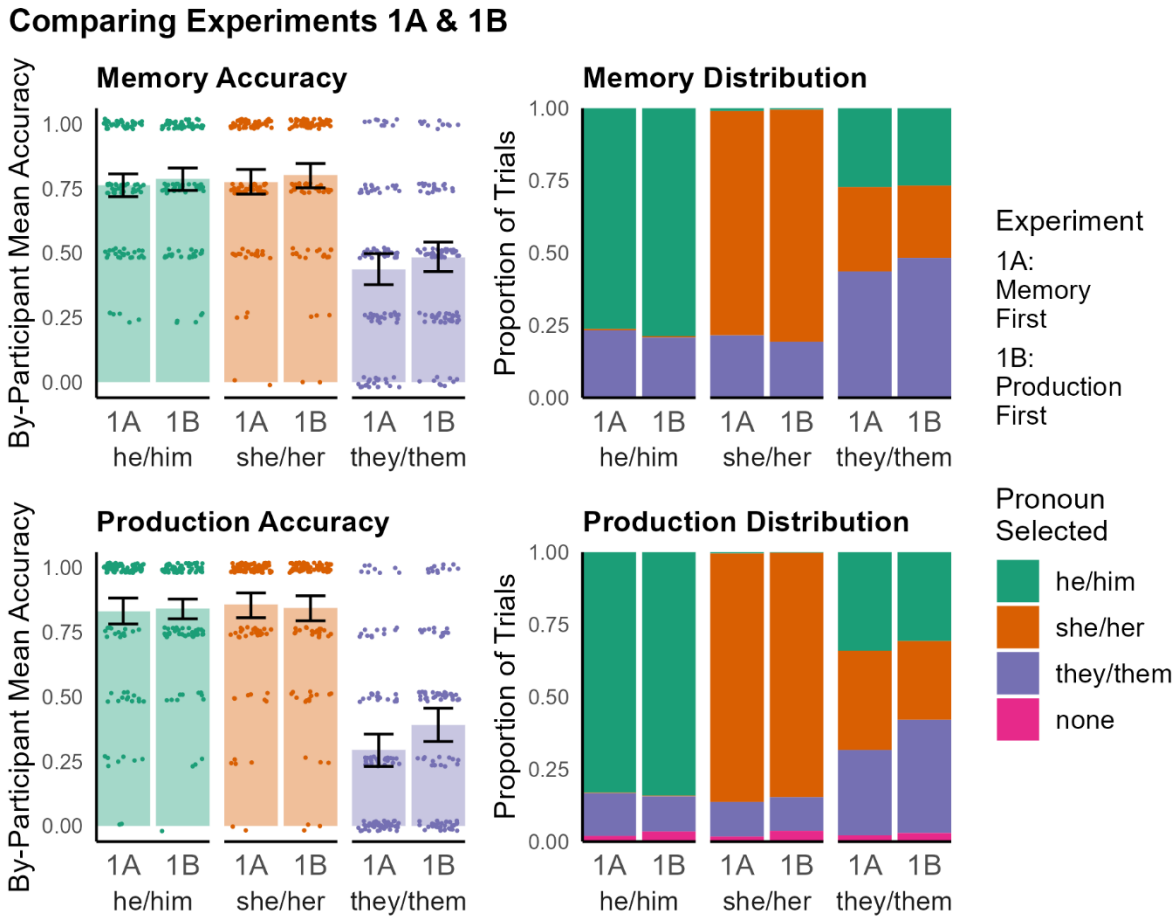


Figure A1.3: Experiments 1A & 1B: Memory and Production. Memory accuracy, distribution of memory responses, production accuracy, and distribution of production responses, comparing between task orders. Points indicate by-participant means, and error bars indicate 95% CIs calculated over the by-participant means.

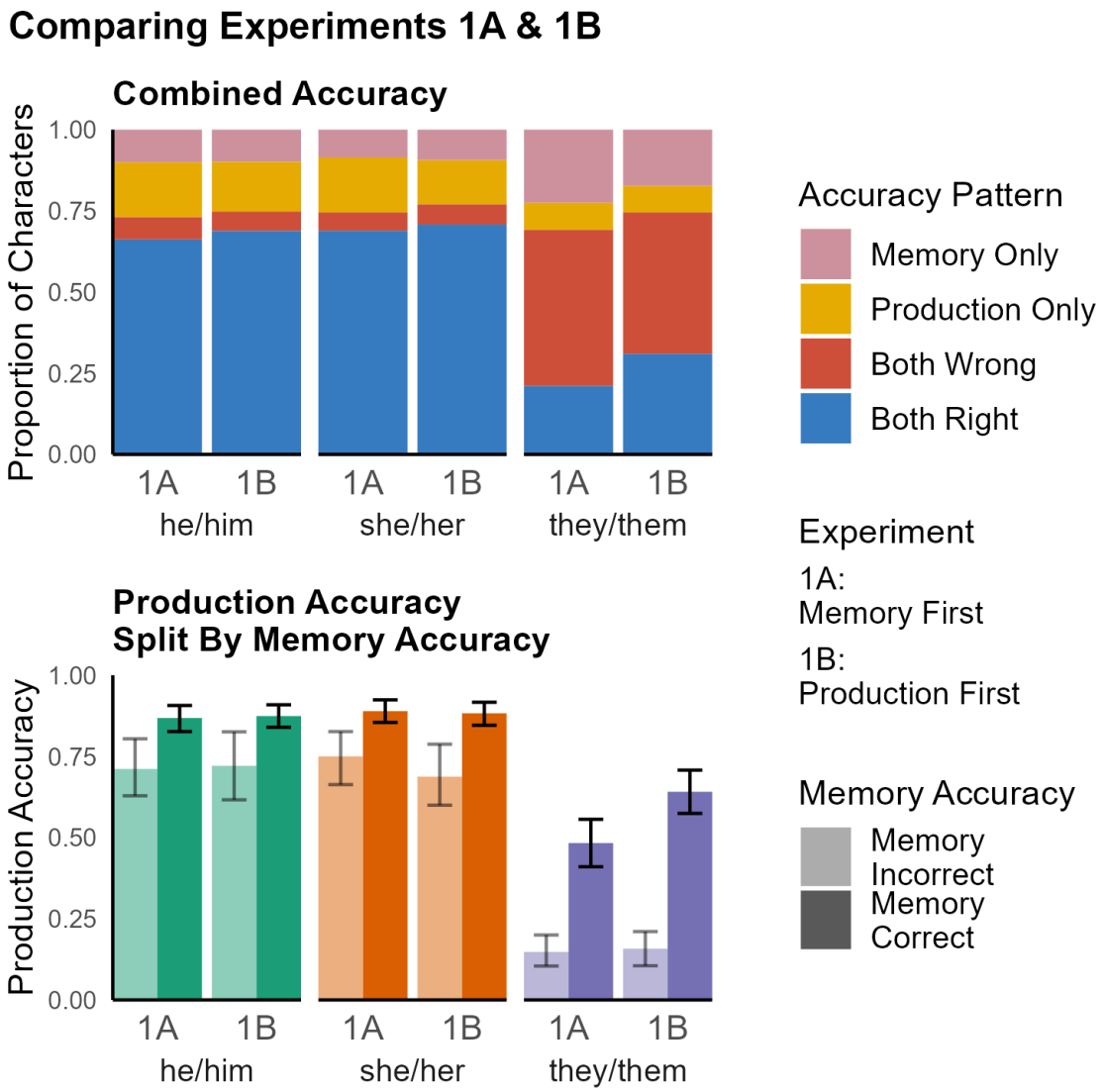


Figure A1.4: Experiments 1A & 1B: Memory Predicting Production. Distribution of combined memory and production accuracy, then production accuracy split by memory accuracy, comparing between task orders. Error bars indicate 95% CIs calculated over trials.

Table A1.7: Experiments 1A & 1B: Memory. Model results for the effects of Pronoun and Task Order on Memory Accuracy.

Comparing Experiments 1A (Memory First) & 1B (Production First): Memory

	Memory Accuracy			
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	0.866	0.070	12.310	**<0.001**
Pronoun: They (-.66) **vs He** (+.33) **+ She** (+.33)	1.582	0.101	15.715	**<0.001**
Pronoun: He (-.5) vs She (+.5)	0.088	0.130	0.674	0.500
Order: Memory First (-.5) vs Production First (+.5)	0.177	0.132	1.341	0.180
Pronoun (They vs He + She) * Order	-0.043	0.196	-0.217	0.828
Pronoun (He vs She) * Order	0.023	0.247	0.094	0.925
Random Effects				
$T_{00\ Participant}$	0.406			
$T_{00\ Name}$	0.005			
$N_{Participant}$	203			
N_{Name}	12			
Observations	2436			

Table A1.8: Experiments 1A & 1B: Production. Model results for the effects of Pronoun and Task Order on Production Accuracy.

Comparing Experiments 1A (Memory First) & 1B (Production First): Production

	Production Accuracy			
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	0.905	0.052	17.425	**<0.001**
Pronoun: They (-.66) **vs He** (+.33) **+ She** (+.33)	2.370	0.102	23.194	**<0.001**
Pronoun: He (-.5) vs She (+.5)	0.112	0.137	0.819	0.413
Order: Memory First (-.5) vs Production First (+.5)	0.134	0.104	1.291	0.197
Pronoun (They vs He + She) * Order	-0.453	0.204	-2.215	**0.027**
Pronoun (He vs She) * Order	-0.187	0.274	-0.683	0.495
Observations	2436			

Table A1.9: Experiments 1A & 1B: Memory Predicting Production. Model results for the effects of Pronoun, Memory Accuracy, and Task Order on Production Accuracy.

Comparing Experiments 1A (Memory First) & 1B (Production First): Memory Predicting Production

	Production Accuracy			
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	0.729	0.056	12.928	**<0.001**
Pronoun: They (-.66) **vs He** (+.33) **+ She** (+.33)	2.208	0.113	19.506	**<0.001**
Pronoun: He (-.5) vs She (+.5)	0.078	0.146	0.536	0.592
Memory Accuracy: Wrong (-.5) **vs Right** (+.5)	1.354	0.113	12.005	**<0.001**
Order: Memory First (-.5) vs Production First (+.5)	0.075	0.113	0.667	0.505
Pronoun (They vs He + She) * Memory Accuracy	-0.932	0.226	-4.115	**<0.001**
Pronoun (He vs She) * Memory Accuracy	0.121	0.293	0.413	0.680
Pronoun (They vs He + She) * Order	-0.435	0.226	-1.923	0.055
Pronoun (He vs She) * Order	-0.238	0.293	-0.813	0.416
Memory Accuracy * Order	0.274	0.226	1.214	0.225
Pronoun (They vs He + She) * Memory Accuracy * Order	-0.448	0.453	-0.989	0.323
Pronoun (He vs She) * Memory Accuracy * Order	0.239	0.585	0.408	0.683
Observations	2436			

Table A1.10: Experiments 1A & 1B: Task Differences. Model results for the effects of Pronoun and Task Order on the difference between memory and production accuracy for each participant.

Comparing Experiments 1A (Memory First) & 1B (Production First): By-Participant Differences Between Memory & Production

	Difference Score			
Predictors	*Estimate*	*SE*	*z*	*p*
(Intercept)	-0.002	0.012	-0.206	0.837
Order: Memory First (-.5) vs Production First (+.5)	0.002	0.024	0.068	0.946
Pronoun: They (-.66) **vs He** (+.33) **+ She** (+.33)	-0.181	0.026	-7.079	**<0.001**
Pronoun: He (-.5) vs She (+.5)	-0.001	0.029	-0.040	0.968
Pronoun (They vs He + She) * Order	0.079	0.051	1.548	0.122
Pronoun (He vs She) * Order	0.027	0.058	0.464	0.643
Observations	609			

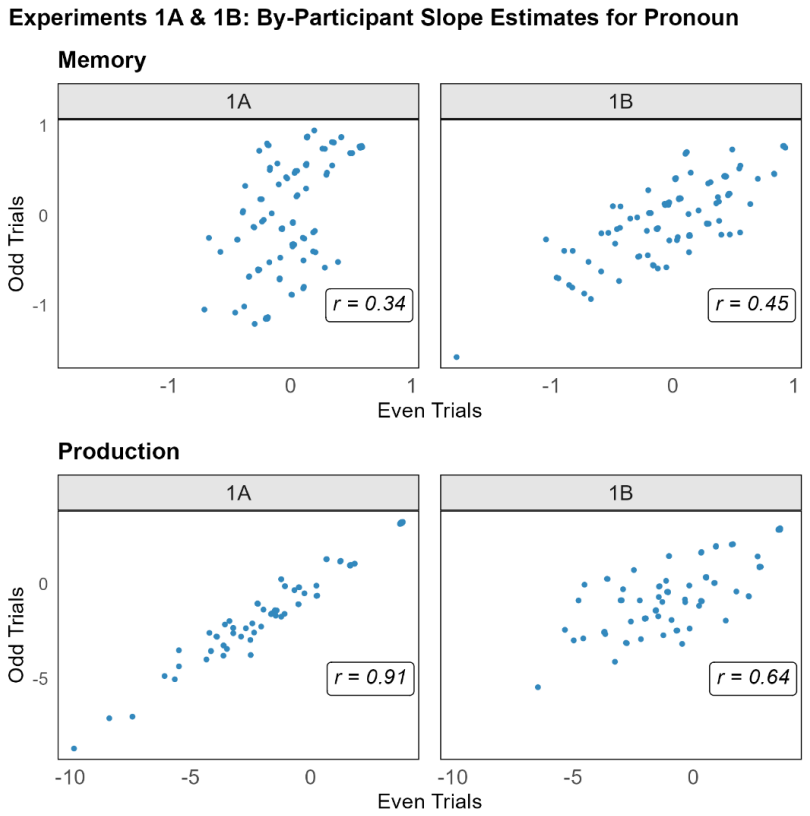


Figure A1.5: Experiments 1A & 1B: Reliability. Correlations between by-participant random slopes for the effect of Pronoun in each half of the data, for the memory and production tasks.

A.2 Experiment 2

A.2.1 Pet & Job Questions

As in Experiments 1A & 1B, memory for the characters' 12 jobs was analyzed in order to make sure the task did not show floor effects, and memory for the characters' 3 pets was analyzed as a less marked comparison to pronouns (Figure A2.1). Averaged across conditions, accuracy for jobs (M = 0.37) was numerically higher than in Experiments 1A (M = 0.21) and 1B (M = 0.29). Accuracy for pets was also higher in Experiment 2 (M = 0.54) than in Experiments 1A (M = 0.41) and 1B (M = 0.43). Job and pet accuracy did not vary based on the characters' pronouns or the PSA and Biography conditions.

Pronoun and pet questions were compared in a model including the Character's Pronouns (contrast coded as in the main analyses), the Question Type (mean-center effects coded), and the PSA and Biography conditions as fixed effects. The initial model included interactions between Character Pronoun, PSA, and Biography as in the main analyses; the interaction between Character Pronoun and Question Type; by-participant and by-item intercepts; and by-participant and by-item slopes for Character Pronoun and Question Type. In addition to the subset of interactions between fixed effects listed above, the most complex model that converged included by-participant intercepts, by-item intercepts, and by-item slopes for Question Type (Table A2.1). Participants were significantly more accurate for pronoun questions than pet questions (β = 1.03, z = 15.59, $p < .001$), and the interaction between Character Pronoun (they/them vs he/him + she/her) and Question Type was significant (β = 1.50, z = 12.84, $p < .001$). Probing this interaction indicated that there was no significant difference in accuracy between pronouns and pets for they/them characters (β = 0.05, z = 0.48, p = .63), only for he/him + she/her characters (β = 1.53, z = 18.57, $p < .001$). This resembles the pattern of results in Experiment 1.

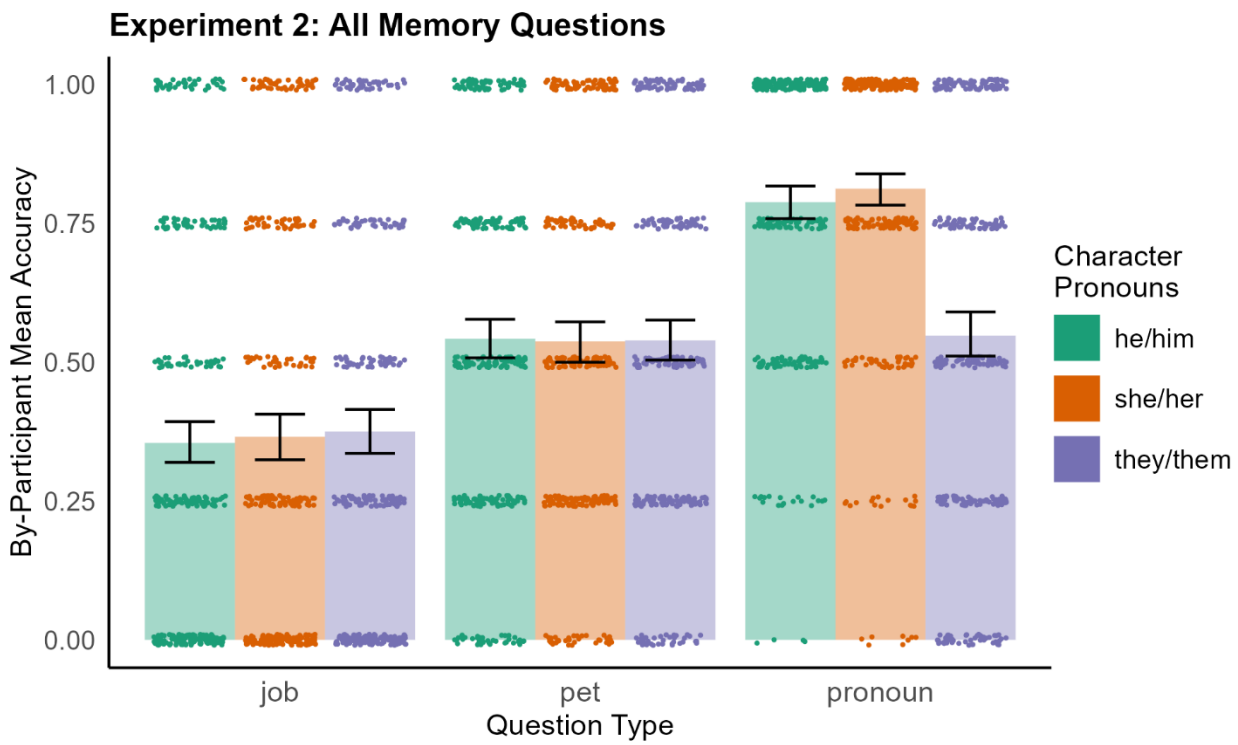


Figure A2.1: Experiment 2: All Memory Questions. Mean accuracy in the multiple-choice memory task for pronouns, pets, and jobs, with colors indicating the character's pronouns. By-participant means are shown as points; error bars indicate 95% CIs calculated over the by-participant means.

Table A2.1: Experiment 2: Comparing Pronoun Accuracy to Pet Accuracy. Model results for the effects of Character Pronoun, PSA, Biography, and Question Type (pronoun vs pet) on Memory Accuracy.

Experiment 2: Memory for Pronouns vs Pets

	Memory Accuracy			
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	0.838	0.089	9.400	**<0.001**
Question Type: Pet (-.5) **vs Pronoun** (+.5)	1.034	0.066	15.592	**<0.001**
Char. Pronoun: They (-.66) **vs He** (+.33) **+ She** (+.33)	0.752	0.058	12.915	**<0.001**
Character Pronoun: He (-.5) vs She (+.5)	0.070	0.087	0.804	0.421
PSA: Unrelated (-.5) vs PSA: Gendered Language (+.5)	0.142	0.165	0.861	0.389
Biography: He/She (-.5) vs Biography: They (+.5)	-0.123	0.165	-0.747	0.455
Question Type * Char. Pronoun (They vs He + She)	1.495	0.116	12.841	**<0.001**
Question Type * Character Pronoun (He vs She)	0.189	0.155	1.220	0.222
Pronoun (They vs He + She) * PSA	-0.247	0.116	-2.135	**0.033**
Pronoun (He vs She) * PSA	-0.026	0.138	-0.188	0.851
PSA * Biography	0.166	0.330	0.503	0.615
Pronoun (They vs He + She) * Biography	-0.011	0.116	-0.095	0.925
Pronoun (He vs She) * Biography	-0.008	0.138	-0.057	0.954
Pronoun (They vs He + She) * PSA * Biography	0.084	0.231	0.363	0.717
Pronoun (They vs He + She) * PSA * Biography	0.357	0.276	1.293	0.196
Random Effects				
$\tau_{00\ Participant}$	1.883			
$\tau_{00\ Name}$	0.011			
$\tau_{11\ Question\ Type\ |\ Participant}$	0.015			
$\rho_{01\ Name}$	0.099			
N_{Name}	12			
$N_{Participant}$	320			
Observations	7680			

A.2.2 Additional Figures

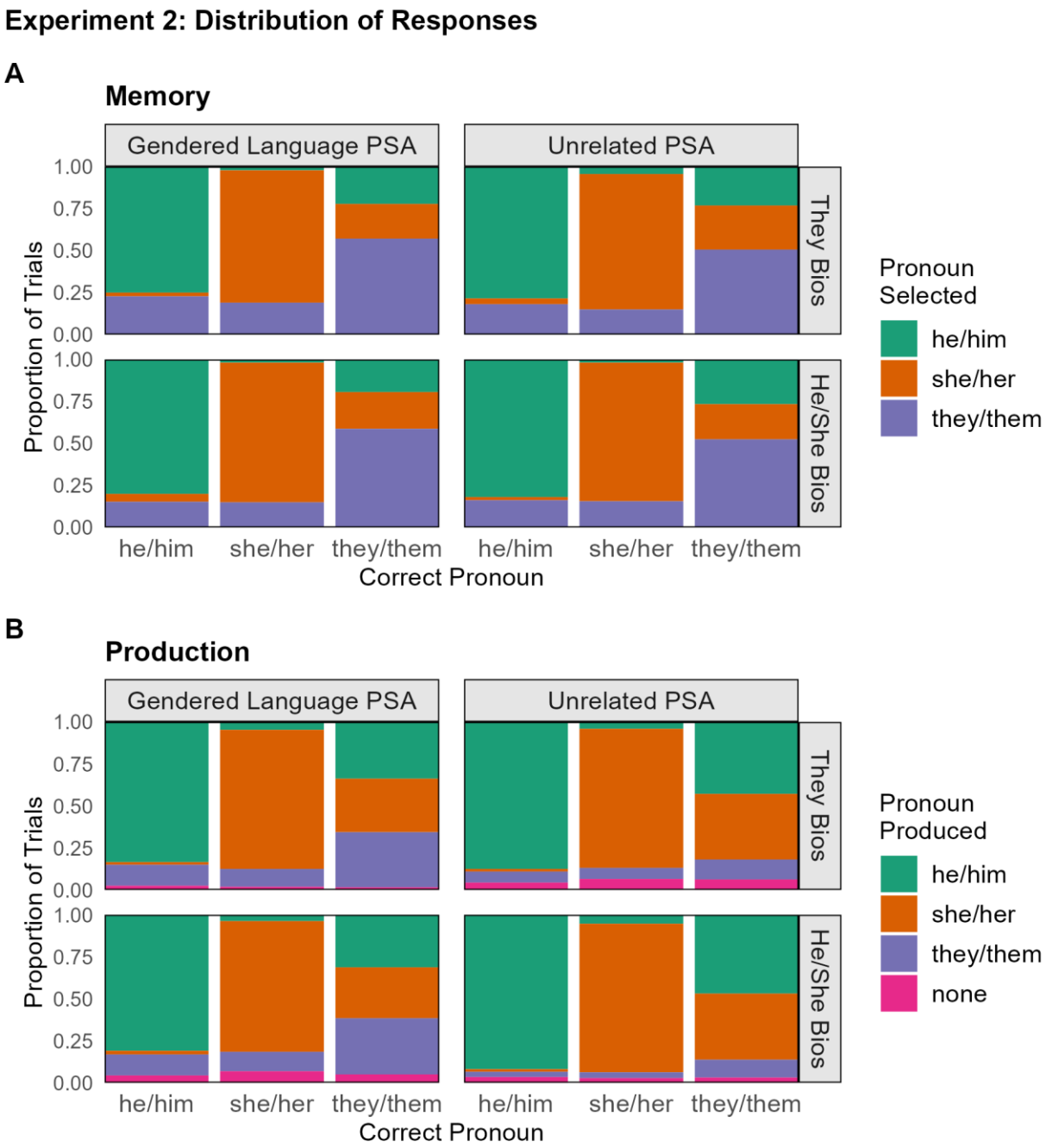


Figure A2.2: Experiment 2: Distribution. Distributions of memory and production responses.

A.2.3 Additional Model Results

Table A2.2: Experiment 2: Memory Predicting Production. Model results for the effects of Pronoun, PSA, Biography, and Memory Accuracy on Production Accuracy.

Experiment 2: Memory Predicting Production

	Production Accuracy			
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	0.522	0.055	9.542	**<0.001**
Pronoun: They (-.66) **vs He** (+.33) **+ She** (+.33)	3.130	0.110	28.527	**<0.001**
Pronoun: He (-.5) **vs She** (+.5)	-0.342	0.133	-2.560	**0.010**
Memory Accuracy: Wrong (-.5) **vs Right** (+.5)	0.830	0.104	8.002	**<0.001**
PSA: Unrelated (-.5) vs PSA: Gendered Language (+.5)	0.075	0.104	0.719	0.472
Biography: He/She (-.5) vs Biography: They (+.5)	-0.017	0.104	-0.163	0.871
Pronoun (They vs He + She) * PSA	-1.896	0.219	-8.677	**<0.001**
Pronoun (He vs She) * PSA	0.120	0.258	0.464	0.643
Pronoun (They vs He + She) * Memory Accuracy	-0.402	0.219	-1.838	0.066
Pronoun (He vs She) * Memory Accuracy	0.255	0.259	0.987	0.324
PSA * Memory Accuracy	0.240	0.207	1.156	0.248
PSA * Biography	0.377	0.207	1.820	0.069
Biography * Memory Accuracy	-0.114	0.207	-0.551	0.582
Pronoun (They vs He + She) * Biography	-0.304	0.219	-1.391	0.164
Pronoun (He vs She) * Biography	0.174	0.258	0.675	0.500
Pronoun (They vs He + She) * PSA * Memory Accuracy	-0.237	0.437	-0.542	0.588
Pronoun (He vs She) * PSA * Memory Accuracy	0.480	0.516	0.931	0.352
PSA * Biography * Memory Accuracy	0.425	0.415	1.025	0.305

Pronoun (They vs He + She) * PSA * Biography	1.198	0.437	2.740	**0.006**
Pronoun (He vs She) * PSA * Biography	0.446	0.516	0.863	0.388
Pronoun (They vs He + She) * Biography * Memory Accuracy	0.830	0.437	1.899	0.058
Pronoun (He vs She) * Biography * Memory Accuracy	-0.390	0.516	-0.756	0.450
Pronoun (They vs He + She) * PSA * Biography * Memory Accuracy	-1.527	0.876	-1.744	0.081
Pronoun (He vs She) * PSA * Biography * Memory Accuracy	-1.203	1.033	-1.164	0.244
Random Effects				
$T_{00\ Name}$	0.004			
N_{Name}	12			
Observations	3840			

Table A2.3: Experiment 2: Produce They*. Model results for the effects of PSA and Biography on whether each participant produced singular* they *at least once. Participants were coded with a 1 if they produced singular* they *at least once in the written sentence completion task, regardless of accuracy, and with a 0 if they did not.*

Experiment 2: Production of Singular *They*

	Produce They/Them			
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	-0.362	0.117	-3.086	**0.002**
PSA: Unrelated (-.5) **vs PSA: Gendered Language** (+.5)	0.927	0.235	3.946	**<0.001**
Biography: He/She (-.5) vs Biography: They (+.5)	0.006	0.235	0.025	0.980
PSA * Biography	-0.816	0.470	-1.737	0.082
Observations	320			

A.3 Experiment 3

A.3.1 Additional Survey Results

Table A3.1: Experiment 3: Additional Demographics. Race/ethnicity has a higher total, as participants could select more than one option.

Additional Participant Demographics	
Education	*162*
Less than high school	2
High school graduate	23
Some college	24
2-year degree	16
4-year degree	63
Professional degree	23
Doctorate	7
Prefer not to answer / Missing data	4
English Experience	*162*
Native (learned from birth)	151
Fully competent in speaking, listening, reading, and writing, but not native	8
Limited but adequate competence in speaking, reading, and writing	0
Restricted ability (e.g., only reading or speaking/listening)	0
Some familiarity (e.g., a year of instruction in school	0
Prefer not to answer / Missing data	3
Race/Ethnicity	*176*
American Indian or Alaska Native	4
Asian	9
Black, African American, or African	21
Hispanic, Latino, or Spanish	19
Middle Eastern or North African	2
Native Hawaiian or Pacific Islander	0
White	112
I use a different term	2
Prefer not to answer / Missing data	7
Total Participants	**162**

The difference in naturalness ratings between singular *they* coreferring with generic and quantified referents [Indefinite] and with masculine, feminine, and gender-neutral names [Proper Name] was tested using a linear mixed-effects model, with Referent Type mean-center effects coded. The model also included by-item intercepts and by-participant slopes. Ratings were mean centered such that a score of 0 indicated the center of the Likert scale, so the significant intercept means that both types of sentences were rated as more natural than unnatural (β = 1.07, t = 10.23, p < .001). The significant effect of Referent Type indicated that singular *they* was rated as more natural with indefinites than with proper names (β = 0.98, t = 4.89, p < .001) (Table A3.2).

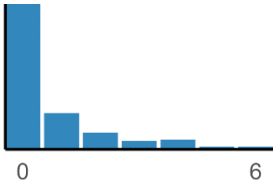
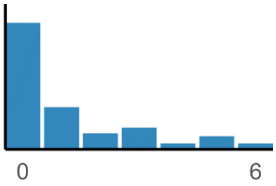
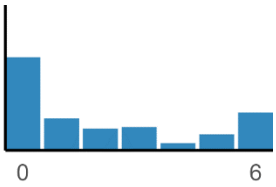
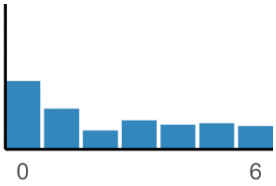
Table A3.2: Experiment 3: Naturalness Ratings. Model results for the effect of Referent Type (generic, each, every vs masculine, feminine, gender-neutral names) on sentence naturalness ratings for singular they.

Experiment 3: Naturalness Ratings for Singular *They*

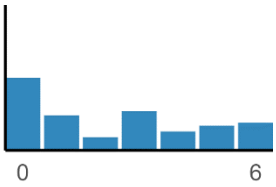
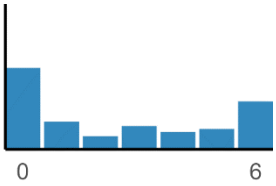
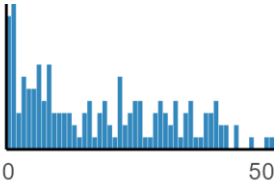
	Naturalness Ratings			
Predictors	*Estimates*	*SE*	*t*	*p*
(Intercept)	1.067	0.104	10.234	**<0.001**
Referent Type: Proper Name (-.5) **vs Indefinite** (+.5)	0.981	0.200	4.893	**<0.001**
Random Effects				
σ^2	1.294			
$\tau_{00\ Participant}$	0.938			
$\tau_{00\ Item}$	0.022			
$\tau_{11\ Referent\ Type\ |\ Participant}$	3.229			
$\rho_{01\ Participant}$	-0.555			
ICC	0.577			
N_{Item}	6			
$N_{Participant}$	159			
Observations	954			
Marginal R^2 / Conditional R^2	0.073	/ 0.608		

Table A3.3: Experiment 3: Gender Beliefs. Question texts, distributions, means, and SDs for items in the gender beliefs scale (Nagoshi et al., 2008).

Experiment 3: Gender Binary & Gender Essentialism Beliefs

Item	*M*	*SD*	*Distribution*
I avoid people on the street whose gender is unclear to me.	0.74	1.32	
I am uncomfortable around people who don't conform to traditional gender roles, e.g., aggressive women or emotional men.	1.14	1.66	
I would be upset if someone I'd known a long time revealed to me that they used to be another gender.	1.40	1.81	
I think there is something wrong with a person who says that they are neither a man nor a woman.	1.97	2.25	
I believe that a person can never change their gender.	2.08	2.29	
When I meet someone, it is important for me to be able to identify them as a man or a woman.	2.31	2.11	

Experiment 3: Gender Binary & Gender Essentialism Beliefs

Item	*M*	*SD*	*Distribution*
I don't like it when someone is flirting with me, and I can't tell if they are a man or a woman.	2.35	2.16	
A person's genitalia define what gender they are, e.g., a penis defines a person as being a man, a vagina defines a person as being a woman.	2.52	2.41	
I believe that the male/female dichotomy is natural.	3.50	2.17	
Total	18.02	14.38	

0: Strongly Disagree – 6: Strongly Agree

A.3.2 Additional Pronoun Accuracy Results

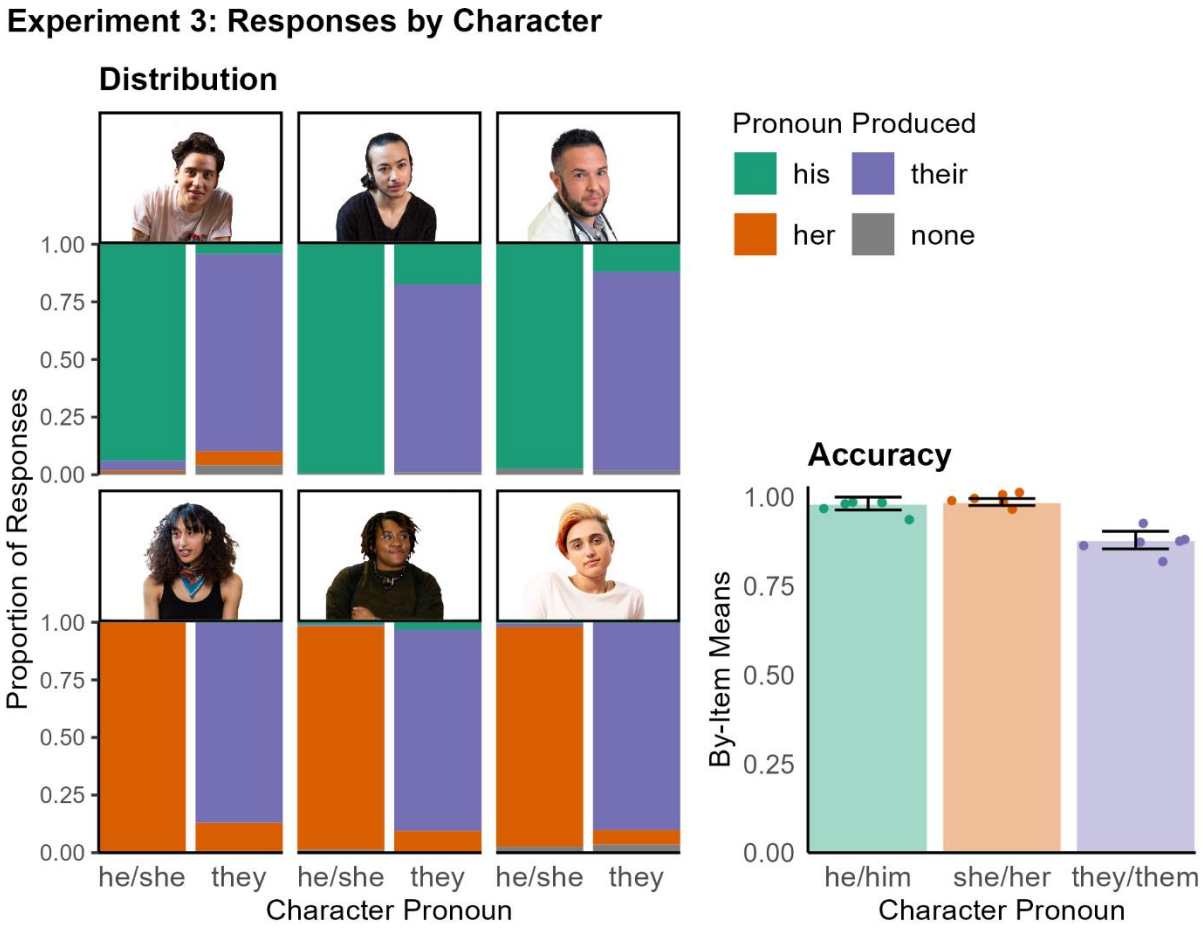


Figure A3.1: Experiment 3: By Character. [A] Distribution of pronouns produced for each of the 6 character images (Drucker, 2019). [B] Mean accuracy for each of the 18 name + image + pronoun combinations, summarized across Nametag & Introduction conditions. Error bars indicate 95% CIs calculated over the by-item means.

A.3.3 Participant Covariate Analyses

Figure A3.2: Experiment 3: Participant Covariate Distributions. Distribution of mean-centered, rescaled participant covariates tested for model inclusion.

Table A3.4: Experiment 3: Participant Covariates. Model results for the effects of Pronoun, Nametag, Introduction, Gender Beliefs, and Familiarity with Pronoun-Sharing on Accuracy.

Experiment 3: Participant Covariates

	Production Accuracy			
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	11.431	1.398	8.175	**<0.001**
Pronoun Pair: T|HS (-.66) **vs HS|T** (+.33) **+ HS|SH** (+.33)	4.009	1.102	3.638	**<0.001**
Pronoun Pair: HS|T (-.5) vs HS|SH (+.5)	-0.900	0.730	-1.233	0.218
Nametag (-.5; +.5)	1.332	1.136	1.173	0.241
Introduction (-.5; +.5)	0.462	1.188	0.389	0.697
Gender Beliefs (mean-centered, higher binary endorsement +)	-10.067	2.927	-3.439	**0.001**
Familiarity with Pronoun Sharing (mean-centered)	-3.232	1.655	-1.953	0.051
Pronoun (T|HS vs HS|T + HS|SH) *** Nametag**	2.734	0.940	2.908	**0.004**
Pronoun (HS|T vs HS|SH) * Nametag	-0.326	1.189	-0.274	0.784
Pronoun (T|HS vs HS|T + HS|SH) *** Intro**	-3.445	1.223	-2.816	**0.005**
Pronoun (HS|T vs HS|SH) * Intro	-1.450	1.416	-1.024	0.306
Nametag * Intro	-1.428	2.286	-0.625	0.532
Pronoun (T|HS vs HS|T + HS|SH) *** Gender Beliefs**	6.504	1.841	3.533	**<0.001**
Pronoun (HS|T vs HS|SH) * Gender Beliefs	-4.167	2.573	-1.620	0.105
Intro * Gender Beliefs	8.753	5.573	1.571	0.116
Pronoun (T|HS vs HS|T + HS|SH) * Familiarity	-2.048	1.127	-1.817	0.069
Pronoun (HS|T vs HS|SH) * Familiarity	1.067	1.452	0.735	0.462
Nametag * Gender Beliefs	3.090	4.881	0.633	0.527
Intro * Familiarity	-0.945	3.261	-0.290	0.772
Pronoun (T|HS vs HS|T + HS|SH) *** Nametag * Intro**	6.163	1.651	3.733	**<0.001**
Pronoun (HS|T vs HS|SH) * Nametag * Intro	0.663	2.248	0.295	0.768

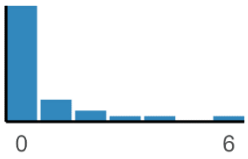
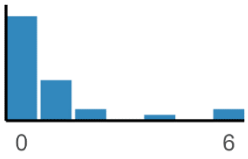
Pronoun (T|HS vs HS|T + HS|SH) *** Gender Beliefs * Intro**	13.319	4.284	3.109	**0.002**
Pronoun (HS|T vs HS|SH) * Gender Beliefs * Intro	2.599	4.359	0.596	0.551
Pronoun (T|HS vs HS|T + HS|SH) * Nametag * Gender Beliefs	-11.034	4.642	-2.377	0.017
Pronoun (HS|T vs HS|SH) * Nametag * Gender Beliefs	-9.794	5.381	-1.820	0.069
Pronoun (T|HS vs HS|T + HS|SH) * Intro * Familiarity	6.643	2.567	2.588	0.010
Pronoun (HS|T vs HS|SH) * Intro * Familiarity	4.790	3.032	1.580	0.114
Random Effects				
$T_{00\ Participant}$	34.933			
$T_{00\ Character}$	3.143			
$N_{Participant}$	158			
$N_{Character}$	18			
Observations	4640			

A.4 Experiment 4

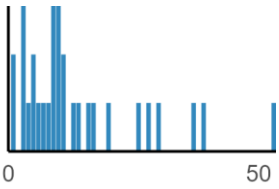
A.4.1 Additional Survey Results

Table A4.1: Experiment 4: Gender Beliefs. Question texts, distributions, means, and SDs for items in the gender beliefs scale (Nagoshi et al., 2008).

Experiment 4: Gender Binary & Gender Essentialism Beliefs

Item	*M*	*SD*	*Distribution*
I avoid people on the street whose gender is unclear to me.	0.67	1.30	
I think there is something wrong with a person who says that they are neither a man nor a woman.	0.70	1.42	
I am uncomfortable around people who don't conform to traditional gender roles, e.g., aggressive women or emotional men.	0.90	1.65	
I believe that a person can never change their gender.	1.03	1.77	
I would be upset if someone I'd known a long time revealed to me that they used to be another gender.	1.37	2.16	

Experiment 4: Gender Binary & Gender Essentialism Beliefs

Item	*M*	*SD*	*Distribution*
A person's genitalia define what gender they are, e.g., a penis defines a person as being a man, a vagina defines a person as being a woman.	1.70	1.95	
When I meet someone, it is important for me to be able to identify them as a man or a woman.	1.87	1.81	
I don't like it when someone is flirting with me, and I can't tell if they are a man or a woman.	2.63	2.14	
I believe that the male/female dichotomy is natural.	3.07	2.00	
Total	13.93	12.56	

0: Strongly Disagree – 6: Strongly Agree

As in Experiment 3, the difference in naturalness ratings between singular *they* coreferring with generic and quantified referents [Indefinite] and with masculine, feminine, and gender-neutral names [Proper Name] was tested using a linear mixed-effects model (Table A4.2). Referent Type was mean-center effects coded, and the model also included by-item intercepts and by-participant slopes. Ratings were mean centered such that a score of 0 indicated the center of the Likert scale, so the significant intercept means that both types of sentences were rated as more natural than unnatural (β = 1.07, t = 3.89, p < .01). There was no significant difference between proper name and indefinite referents (β = 0.19, t = 0.36, p = .73).

Table A4.2: Experiment 4: Naturalness Ratings. Model results for the effect of Referent Type (generic, each, every vs masculine, feminine, gender-neutral names) on sentence naturalness ratings for singular they.

Experiment 4: Naturalness Ratings for Singular *They*

	Naturalness Ratings			
Predictors	*Estimates*	*SE*	*t*	*p*
(Intercept)	1.072	0.276	3.892	**0.004**
Referent Type: Proper Name (-.5) vs Indefinite (+.5)	0.189	0.519	0.364	0.726
Random Effects				
σ^2	1.420			
$\tau_{00\ Participant}$	0.807			
$\tau_{00\ Item}$	0.247			
$\tau_{11\ Referent\ Type\ |\ Participant}$	2.200			
$\rho_{01\ Participant}$	-0.317			
ICC	0.530			
N_{Item}	6			
$N_{Participant}$	30			
Observations	180			
Marginal R^2 / Conditional R^2	0.003 / 0.532			

A.4.2 Match Judgments

Table A4.3: Experiment 4: Test Trial Match Rates. Model results for the effect of Pronoun Pair on the likelihood of judging the story to match the scene in test trials (1 = match = correct).

Experiment 4: Match Judgment Rates in Test Trials

		Match		
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	3.087	0.254	12.176	**<0.001**
Pronoun Pair: They|HeShe (-.66) vs HeShe|They (+.33) + HeShe|SheHe (+.33)	0.330	0.245	1.345	0.179
Pronoun Pair: HeShe|They (-.5) vs HeShe|SheHe (+.5)	0.255	0.275	0.926	0.355
Random Effects				
$\tau_{00\ Participant}$	1.536			
$\tau_{11\ Pronoun\ Pair\ (T\|HS\ vs\ HS\|T\ +\ HS\|SH)\ |\ Participant}$	0.341			
$\tau_{11\ Pronoun\ Pair\ (HS\|T\ vs\ HS\|SH)\ |\ Participant}$	0.120			
$\rho_{01\ Pronoun\ Pair\ (T\|HS\ vs\ HS\|T\ +\ HS\|SH)\ |\ Participant}$	0.277			
$\rho_{01\ Pronoun\ Pair\ (HS\|T\ vs\ HS\|SH)\ |\ Participant}$	-0.637			
$N_{\ Participant}$	30			
Observations	2877			

Table A4.4: Experiment 4: Wrong Pronoun Trial Match Rates. Model results for the effect of Pronoun on judging the story to match the scene in wrong pronoun trials (1 = match = incorrect). There were no wrong pronoun trials for the HeShe|They condition, so the They|HeShe (they/them character referred to with he/him or she/her, whichever the competitor did not use) and the HeShe|SheHe (he/him and she/her characters referred to with they/them) conditions were mean-center effects coded.

Experiment 4: Match Judgment Rates in Wrong Pronoun Trials

		Match		
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	-0.373	0.328	-1.137	0.255
Correct Pronoun: They (-.5) vs He + She (+.5)	-0.076	0.362	-0.211	0.833
Random Effects				
$\tau_{00\ Story}$	0.190			
$\tau_{00\ Participant}$	2.206			
$\tau_{11\ Pronoun\ |\ Story}$	0.140			
$\rho_{01\ Pronoun\ |\ Story}$	-0.507			
$N_{Participant}$	30			
N_{Story}	35			
Observations	238			

Table A4.5: Experiment 4: Test Trial Match RT. Model results for the effect of Pronoun Pair on reaction times for test trial match judgments.

Experiment 4: Match Judgment RT in Test Trials

		RT (ms)		
Predictors	*Estimate*	*SE*	*t*	*p*
(Intercept)	3888.739	32.056	121.311	**<0.001**
Pronoun Pair: They|HeShe (-.66) **vs HeShe|They** (+.33) **+ HeShe|SheHe** (+.33)	-506.668	40.874	-12.396	**<0.001**
Pronoun Pair: HeShe|They (-.5) vs HeShe|SheHe (+.5)	32.961	43.864	0.751	0.452
Random Effects				
σ^2	0.014			
$\tau_{00\ Story}$	126			
$\tau_{00\ Participant}$	305041			
$\tau_{11\ Pronoun\ Pair\ (T\|HS\ vs\ HS\|T\ +\ HS\|SH)\ |\ Story}$	397848			
$\tau_{11\ Pronoun\ Pair\ (HS\|T\ vs\ HS\|SH)\ |\ Story}$	592385			
$\tau_{11\ Pronoun\ Pair\ (T\|HS\ vs\ HS\|T\ +\ HS\|SH)\ |\ Participant}$	114829			
$\tau_{11\ Pronoun\ Pair\ (HS\|T\ vs\ HS\|SH)\ |\ Participant}$	151199			
$\rho_{01\ Pronoun\ Pair\ (T\|HS\ vs\ HS\|T\ +\ HS\|SH)\ |\ Story}$	-0.324			
$\rho_{01\ Pronoun\ Pair\ (HS\|T\ vs\ HS\|SH)\ |\ Story}$	-0.961			
$\rho_{01\ Pronoun\ Pair\ (T\|HS\ vs\ HS\|T\ +\ HS\|SH)\ |\ Participant}$	-0.833			
$\rho_{01\ Pronoun\ Pair\ (HS\|T\ vs\ HS\|SH)\ |\ Participant}$	0.092			
$N_{\ Participant}$	30			
$N_{\ Story}$	60			
Observations	2842			

A.4.3 Additional Figures

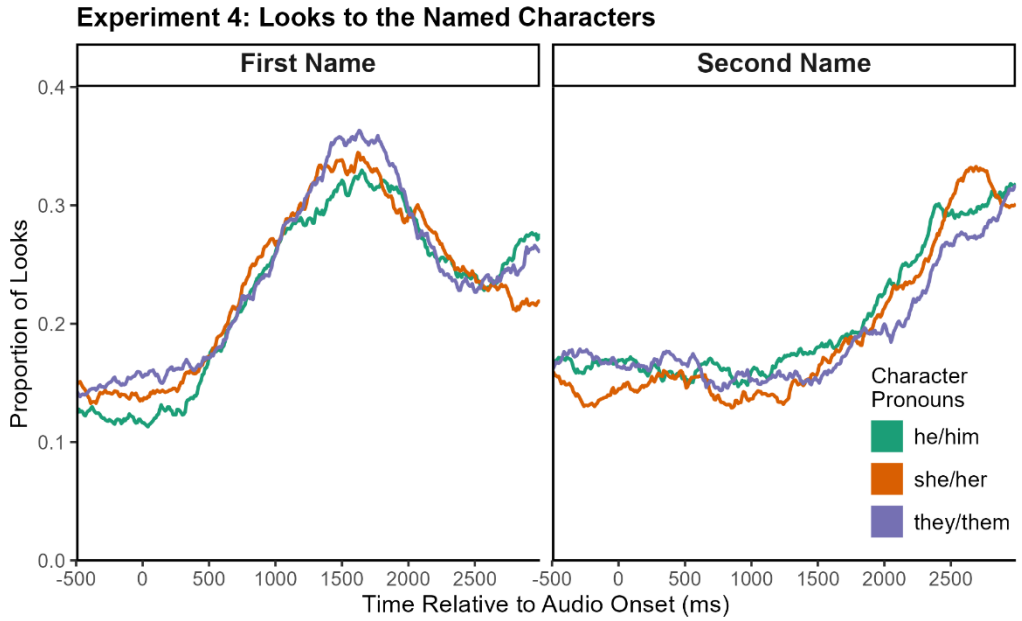


Figure A4.1: Experiment 4: Name Window. Looks to the named characters at the beginning of the story, with the first name beginning at 0ms and the second name beginning ~1000ms later.

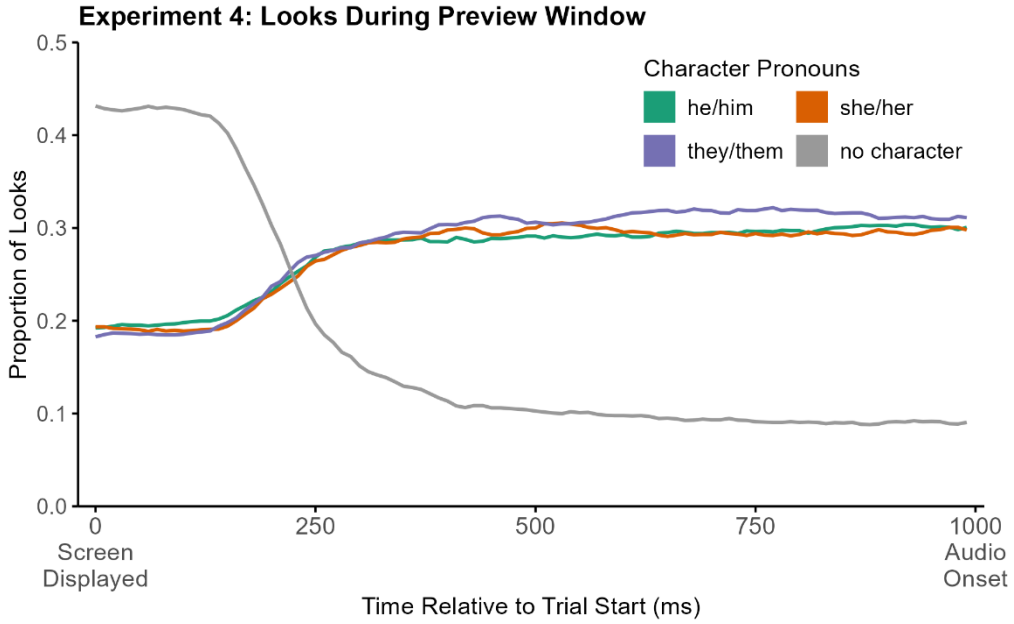


Figure A4.2: Experiment 4: Preview Window. Looks to each character during the 1000ms preview time before audio started, split by the character's pronouns.

A.4.4 Additional Eyetracking Results

The Order of Mention effect was estimated separately for each Pronoun Pair condition by running three models, each with one Pronoun Pair condition coded as 0 and the other two Pronoun Pair conditions coded as 1. The maximal random effects structures that converged (Bates et al., 2015; R Core Team, 2023; Voeten, 2023) only included a subset of those in the main model (Table 4.2). For the by-participant effects, each kept all 3 slopes (AR(1), Order, and Trial Number). For the by-item effects, the HeShe|They reference model included both slopes (Order, Trend), the HeShe|SheHe reference model only included slopes for Trend, and the They|HeShe reference model only included intercepts.

In HeShe|SheHe trials (Table A4.6), neither the main effect of Order (β = 0.05, z = 0.47, p = .64) nor its interaction with Pronoun Pair (β = 0.07, z = 0.70, p = .49) were significant. In HeShe|They trials (Table A4.7), the main effect of Order was significant as anticipated, with participants more likely to be looking at target characters who had been named first in the story than target characters who had been named second (β = 0.22, z = 2.20, p < .05). The interaction between Order and Pronoun Pair was also significant (β = -0.20, z = -2.21, p < .05). Probing this interaction indicated that listeners were less likely to be looking at the target in HeShe|They trials than in HeShe|SheHe and They|HeShe trials when the target was mentioned second (β = 0.17, z = 2.51, p < .05), but not when the target was mentioned first (β = -0.04, z = -0.62, p = .54). In They|HeShe trials (Table A4.8), neither the main effect of Order (β = 0.00, z = -0.03, p = .97) nor its interaction with Pronoun Pair (β = 0.14, z = 1.48, p = .14) were significant.

Table A4.6: Experiment 4: Order in HeShe|SheHe. Model results for the effects of Pronoun Pair and Order on the likelihood of looking at the target character (=1) or not (=0). The HeShe|SheHe condition is coded as 0, and the HeShe|They and They|HeShe conditions are coded as 1.

Experiment 4: Looks to the Target Character (HeShe|SheHe as Reference Group)

	Looks to Target			
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	-4.480	0.089	-50.111	**<0.001**
AR(1) (0, 1)	9.971	0.109	91.859	**<0.001**
Pronoun Pair: HeShe|SheHe (0) vs HeShe|They (1) + They|HeShe (1)	-0.347	0.047	-7.454	**<0.001**
Order: Target Mentioned Second (-.5) vs First (+.5)	0.046	0.098	0.467	0.641
Trend (mean-centered)	0.049	0.038	1.309	0.191
Pronoun Pair * Order	0.065	0.093	0.698	0.485
Random Effects				
$T_{00\ Story}$	0.005			
$T_{00\ Participant}$	0.196			
$T_{11\ Trend\ |\ Story}$	0.004			
$T_{11\ AR(1)\ |\ Participant}$	0.268			
$T_{11\ Order\ |\ Participant}$	0.110			
$T_{11\ Trial\ Number\ |\ Participant}$	0.027			
$\rho_{01\ Trend\ |\ Story}$	0.018			
$\rho_{01\ AR(1)\ |\ Participant}$	-0.296			
$\rho_{01\ Order\ |\ Participant}$	0.084			
$\rho_{01\ Trial\ Number|\ Participant}$	0.508			
$N_{Participant}$	30			
N_{Story}	60			
Observations	296022			

Table A4.7: Experiment 4: Order in HeShe|They. Model results for the effects of Pronoun Pair and Order on the likelihood of looking at the target character (=1) or not (=0). The HeShe|They condition is coded as 0, and the HeShe|SheHe and They|HeShe conditions are coded as 1.

Experiment 4: Looks to the Target Character (HeShe|They as Reference Group)

	Looks to Target			
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	-4.768	0.088	-54.437	**<0.001**
AR(1)	9.988	0.110	90.880	**<0.001**
Pronoun Pair: HeShe|They (0) vs HeShe|SheHe (1) + They|HeShe (1)	0.062	0.046	1.329	0.184
Order: Target Mentioned Second (-.5) **vs First** (+.5)	0.218	0.099	2.196	**0.028**
Trend (mean-centered)	0.044	0.038	1.152	0.249
Pronoun Pair * Order	-0.204	0.093	-2.205	**0.027**
Random Effects				
$T_{00\ Story}$	0.006			
$T_{00\ Participant}$	0.189			
$T_{11\ Order\ |\ Story}$	0.025			
$T_{11\ Trend\ |\ Story}$	0.005			
$T_{11\ AR(1)\ |\ Participant}$	0.274			
$T_{11\ Order\ |\ Participant}$	0.106			
$T_{11\ Trial\ Number\ |\ Participant}$	0.032			
$\rho_{01\ Order\ |\ Story}$	0.951			
$\rho_{01\ Trend\ |\ Story}$	0.130			
$\rho_{01\ AR(1)\ |\ Participant}$	-0.289			
$\rho_{01\ Order\ |\ Participant}$	0.097			
$\rho_{01\ Trial\ Number\ |\ Participant}$	0.544			
$N_{Participant}$	30			
N_{Story}	60			
Observations	296022			

Table A4.8: Experiment 4: Order in They|HeShe. Model results for the effects of Pronoun Pair and Order on the likelihood of looking at the target character (=1) or not (=0). The They|HeShe condition is coded as 0, and the HeShe|They and HeShe|SheHe conditions are coded as 1.

Experiment 4: Looks to the Target Character (They|HeShe as Reference Group)

	Looks to Target			
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	-4.901	0.089	-54.975	**<0.001**
AR(1) (0, 1)	9.983	0.110	90.621	**<0.001**
Pronoun Pair: They|HeShe (0) vs HeShe|SheHe (1) + HeShe|They (1)	0.273	0.046	5.884	**<0.001**
Order: Target Mentioned Second (-.5) vs First (+.5)	-0.003	0.099	-0.034	0.973
Trend (mean-centered)	0.047	0.037	1.269	0.205
Pronoun Pair * Order	0.138	0.093	1.479	0.139
Random Effects				
$T_{00\ Story}$	0.005			
$T_{00\ Participant}$	0.192			
$T_{11\ AR(1)\ |\ Participant}$	0.277			
$T_{11\ Order\ |\ Participant}$	0.110			
$T_{11\ Trial\ Number\ |\ Participant}$	0.030			
$\rho_{01\ AR(1)\ |\ Participant}$	-0.301			
$\rho_{01\ Order\ |\ Participant}$	0.101			
$\rho_{01\ Trial\ Number\ |\ Participant}$	0.493			
$N_{Participant}$	30			
N_{Story}	60			
Observations	296022			

Table A4.9: Experiment 4: Target Pronoun. Model results for the effects of Target Pronoun and Order on the likelihood of looks to the target character (=1) or not (=0), during the window starting 200ms after pronoun onset and ending at 1210ms, the earliest shape word onset.

Experiment 4: Target Pronoun

	Looks to Target			
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	-4.723	0.084	-56.372	**<0.001**
AR(1) (0, 1)	9.999	0.108	92.291	**<0.001**
Target Pronoun: They (-.66) **vs He** (+.33) **+ She** (+.33)	0.275	0.047	5.864	**<0.001**
Target Pronoun: She (-.5) vs He (+.5)	0.039	0.053	0.728	0.467
Order: Target Mentioned Second (-.5) vs First (+.5)	0.085	0.077	1.099	0.272
Trend (mean-centered)	0.047	0.037	1.291	0.197
Target Pronoun (They vs He + She) * Order	0.139	0.094	1.478	0.140
Target Pronoun (She vs He) * Order	0.200	0.107	1.867	0.062
Random Effects				
$T_{00\ Story}$	0.005			
$T_{00\ Participant}$	0.191			
$T_{11\ AR(1)\ |\ Participant}$	0.278			
$T_{11\ Order\ |\ Participant}$	0.111			
$T_{11\ Trial\ Number\ |\ Participant}$	0.029			
$T_{11\ Order\ *\ Trial\ Number\ |\ Participant}$	0.014			
$\rho_{01\ AR(1)\ |\ Participant}$	-0.300			
$\rho_{01\ Order\ |\ Participant}$	0.103			
$\rho_{01\ Trial\ Number\ |\ Participant}$	0.488			
$\rho_{01\ Order\ *\ Trial\ Number\ |\ Participant}$	-0.339			
$N_{Participant}$	30			
N_{Story}	60			
Observations	296022			

Table A4.10: Experiment 4: Trend Interactions. Model results for the interactions between Trend, Pronoun Pair, and Order on the likelihood of looking at the target character (=1) or not (=0).

Experiment 4: Interactions with Trend

	Looks to Target			
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	-4.711	0.083	-56.631	**<0.001**
AR(1) (0, 1)	9.974	0.109	91.708	**<0.001**
Pronoun Pair: T|HS (-.66) **vs HS|T** (+.33)**+ H|SH** (+.33)	0.284	0.061	4.624	**<0.001**
Pronoun Pair: HS|T (-.5) **vs HS|SH** (+.5)	0.264	0.073	3.594	**<0.001**
Order: Target Mentioned Second (-.5) vs First (+.5)	0.085	0.078	1.082	0.279
Trend (mean-centered)	0.058	0.037	1.587	0.113
Pronoun Pair (T|HS vs HS|T + HS|SH) * Order	0.132	0.094	1.403	0.161
Pronoun Pair (HS|T vs HS|SH) * Order	-0.188	0.107	-1.755	0.079
Pronoun Pair (T|HS vs HS|T + HS|SH) * Trend	-0.109	0.079	-1.392	0.164
Pronoun Pair (HS|T vs HS|SH) * Trend	-0.135	0.089	-1.519	0.129
Order * Trend	0.032	0.073	0.438	0.661
Pronoun Pair (T|HS vs HS|T + HS|SH) * Order * Trend	-0.165	0.157	-1.054	0.292
Pronoun Pair (HS|T vs HS|SH) * Order * Trend	-0.115	0.176	-0.654	0.513
Random Effects				
T_{00} Participant	0.193			
T_{11} AR(1) | Participant	0.268			
T_{11} Order | Participant	0.118			
T_{11} Trial Number | Participant	0.030			
T_{11} Pronoun Pair (T|HS vs HS|T + HS|SH) | Participant	0.038			
T_{11} Pronoun Pair (HS|T vs HS|SH) | Participant	0.065			
ρ_{01} AR(1) | Participant	-0.285			
ρ_{01} Order | Participant	0.063			
ρ_{01} Trial Number | Participant	0.463			
ρ_{01} Pronoun Pair (T|HS vs HS|T + HS|SH) | Participant	0.441			
ρ_{01} Pronoun Pair (HS|T vs HS|SH) | Participant	0.272			
N Participant	30			
Observations	296022			

Table A4.11: Experiment 4: AR(1) Interactions. Model results for the interactions between AR(1), Pronoun Pair, and Order on the likelihood of looking at the target character (=1) or not (=0).

Experiment 4: Interactions with AR(1)

	Looks to Target			
Predictors	*Log-Odds*	*SE*	*z*	*p*
(Intercept)	-4.708	0.085	-55.706	**<0.001**
AR(1)	9.972	0.109	91.872	**<0.001**
Pronoun Pair: T|HS (-.66) **vs HS|T** (+.33) **+ HS|SH** (+.33)	0.322	0.060	5.378	**<0.001**
Pronoun Pair: HS|T (-.5) **vs HS|SH** (+.5)	0.276	0.066	4.170	**<0.001**
Order: Target Mentioned Second (-.5) vs First (+.5)	0.081	0.083	0.976	0.329
Trend (mean-centered)	0.052	0.037	1.409	0.159
Pronoun Pair (T|HS vs HS|T + HS|SH) * Order	0.070	0.120	0.582	0.561
Pronoun Pair (HS|They vs HS|SH) * Order	-0.016	0.132	-0.119	0.906
Pronoun Pair (T|HS vs HS|T + HS|SH) * AR(1)	-0.111	0.098	-1.131	0.258
Pronoun Pair (HS|T vs HS|SH) * AR(1)	-0.002	0.112	-0.014	0.989
AR(1) * Order	0.022	0.093	0.233	0.815
Pronoun Pair (T|HS vs HS|T + HS|SH) * Order * AR(1)	0.179	0.195	0.917	0.359
Pronoun Pair (HS|T vs HS|SH) * Order * AR(1)	-0.447	0.223	-2.001	0.045
Random Effects				
T_{00} Story	0.006			
T_{00} Participant	0.197			
T_{11} Trial Number | Story	0.012			
T_{11} AR(1) | Participant	0.270			
T_{11} Order | Participant	0.113			
T_{11} Trial Number | Participant	0.027			
ρ_{01} Trial Number | Story	0.048			
ρ_{01} AR(1) | Participant	-0.293			
ρ_{01} Order | Participant	0.088			
ρ_{01} Trial Number | Participant	0.485			
N Participant	30			
N Story	60			
Observations	296022			

Appendix B

Materials

B.1 Demographics Questions

B.1.1 Experiments 1 & 2

Age: _____

What is your gender? _____

Please rate your overall ability in the English language:

- ○ Native (learned from birth)
- ○ Fully competent in speaking, listening, reading, and writing, but not native
- ○ Limited but adequate competence in speaking, reading, and writing
- ○ Restricted ability (e.g., only reading or speaking/listening)
- ○ Some familiarity (e.g., a year of instruction in school)
- ○ Prefer not to answer

B.1.2 Experiments 3 & 4

Age: _____

What is your gender? _____
(select all that apply)

- ☐ My gender is the same what was written on my original birth certificate
- ☐ My gender is different than what was written on my original birth certificate
- ☐ I consider myself cisgender
- ☐ I consider myself transgender
- ☐ I don't consider myself cisgender or transgender
- ☐ Prefer not to answer

How do you describe your sexuality?
(select all that apply)

- ☐ Asexual
- ☐ Bisexual/Pansexual
- ☐ Gay/Lesbian
- ☐ Heterosexual/Straight
- ☐ Queer
- ☐ Questioning
- ☐ Prefer not to answer
- ☐ I use a different term: _____

How do you describe your race/ethnicity?
(select all that apply)

- ☐ American Indian or Alaska Native
- ☐ Asian
- ☐ Black, African American, or African
- ☐ Hispanic, Latino, or Spanish,
- ☐ Middle Eastern or North African
- ☐ Native Hawaiian or Pacific Islander
- ☐ White
- ☐ Prefer not to answer
- ☐ I use a different term: _____

What is your highest education level?

- ○ Less than high school
- ○ High school graduate
- ○ Some college
- ○ 2-year degree
- ○ 4-year degree
- ○ Professional degree
- ○ Doctorate
- ○ Prefer not to answer

Please rate your overall ability in the English language:

- ○ Native (learned from birth)
- ○ Fully competent in speaking, listening, reading, and writing, but not native
- ○ Limited but adequate competence in speaking, reading, and writing
- ○ Restricted ability (e.g., only reading or speaking/listening)
- ○ Some familiarity (e.g., a year of instruction in school)
- ○ Prefer not to answer

B.2 PSAs (Experiment 2)

Pronoun Guide

This guide is created to help anyone learn how to use people's correct pronouns. Everyone in your community should engage in learning, educating, and advocating for the inclusive use of pronouns for all.

What Are Pronouns?

Pronouns are the words you may like others to use for you in place of your proper name. Some examples include "she/her" or "he/him" or gender-neutral pronouns, such as "they/them." Here is an example of using "they/them" in a sentence: John is substituting for me today and they are an incredible mathematician. Singular "they" pronouns have been used in the English language for centuries. If you are still struggling with using this, remember, it takes practice!

Why Focus on Pronouns?

You may have noticed that people are sharing their pronouns in introductions, on name tags, and at the beginning of meetings. This gives everyone in the room the opportunity to self-identify instead of assuming someone's identity or which pronouns they use. Including pronouns is a first step toward respecting people's identity and creating a more welcoming space for people of all genders.

How Is This More Inclusive?

Pronouns can sometimes be a signifier for someone's gender identity, but not always. We do not want to assume people's gender identity based on expression (typically shown through clothing, hairstyle, mannerisms etc.) By providing an opportunity for people to share their pronouns, you're showing that you're not assuming that their gender identity is based on their appearance.

Mistakes and Misgendering

Misgendering refers to the experience of being labeled by others as a gender other than one that a person identifies with. Because many (not all) associate their pronouns with their gender identity, using the wrong pronouns intentionally or unintentionally is a form of misgendering. If you accidentally use the wrong pronoun when identifying someone, please apologize or say, "thank you," and immediately use the right pronoun,

> *i.e., "This is Alex, she is one of my science students."*
>
> *(you are corrected because Alex uses they/them/theirs pronouns).*
>
> *"Sorry, they are one of my science students" or "Thank you, they are one of my science students."*

Everyone makes mistakes; please take accountability for your mistake and start using the correct pronoun. The important thing to be mindful of is to not unload your guilty feelings on transgender, nonbinary, and gender nonconforming people or expect forgiveness. They might have a strong reaction to the misuse of their pronouns and need space to recenter themselves.

Practice, practice, practice! Use gender-neutral pronouns such as “they” while visualizing the person who uses them. This is especially useful to do right before you’re about to see the person.

When addressing groups of people or people whose pronouns you haven’t been told, use gender-neutral language such as, “siblings,” “third graders,” “students,” “friends,” “folks,” “all,” or “y’all,” rather than “brothers and sisters,” or “guys,” “ladies,” “ma’am,” or “sir.”

Use descriptive language if you do not know a person’s gender, pronouns, or name. e.g., “Can you give this paper to the person across the room with the white t-shirt and short brown hair?”

This text is adapted from the 2020 GLSEN Pronoun Guide: https://www.glsen.org/activity/pronouns-guide-glsen

The choice to spay or neuter your pet may be one of the most important decisions you make impacting their long-term health—and your wallet!

Your pet's health and longevity

The average lifespan of spayed and neutered cats and dogs is demonstrably longer than the lifespan of those not. A study conducted by Banfield Pet Hospitals on a database of 2.2 million dogs and 460,000 cats concluded that neutered male dogs lived 18% longer, and spayed female dogs lived 23% longer. Spayed female cats in the study lived 39% longer, and neutered male cats lived 62% longer. The reduced lifespan of unaltered pets can, in part, be attributed to an increased urge to roam (exposing them to fights with other animals resulting in injuries and infections), to trauma from vehicle strikes, and to other accidental mishaps. Another contributor to the increased longevity of altered pets is their reduced risk of certain types of cancers.

Curb unwanted behaviors

Intact dogs are more prone to urine-marking than neutered dogs. Although urine-marking is usually associated with male dogs, females may do it too. Spaying or neutering your dog should reduce urine-marking and may even stop it altogether.

For cats, the urge to spray is extremely strong in those not altered, so the simplest solution is to alter by 5 months old before the problem arises. Neutering solves 90% of all marking issues, even in cats that have been doing it for a while. It can also minimize howling, the urge to roam and fighting with other males.

In both cats and dogs, the longer you wait, the greater the risk of the surgery not doing the trick because the animal has practiced the behavior for a longer period of time, thereby reinforcing the habit.

Other behavioral problems that can be alleviated by spay/neuter include:

- Roaming, especially when females are in heat
- Excessive barking and mounting activity

While having your pets spayed/neutered can help curb undesirable behaviors, it will not change their fundamental personalities.

Cut costs

When you consider the potential long-term medical costs incurred for an unaltered pet, the savings afforded by spay/neuter are clear, especially given the plethora of low-cost spay/neuter clinics now available.

Caring for a pet with reproductive system cancer or pyometra can easily run thousands of dollars—five to 10 times as much as a routine spay or neuter surgery. In cases where intact dogs and cats may fight, treatment of their related injuries can also result in high veterinary costs.

Spread the word

By spaying or neutering your pet, you can help protect them against certain illnesses, help address unwanted behaviors, save money, and save lives by reducing overpopulation. Contact your veterinarian or your local low-cost spay/neuter clinic to discuss and/or schedule your pet's appointment, and encourage your family and friends to do the same. If your community does not have affordable, accessible spay/neuter and wellness services available, advocate for this funding so that all pet owners have local access to quality medical care for their pets.

This text is adapted from The Humane Society of the United States: https://www.humanesociety.org/resources/why-you-should-spayneuter-your-pet

B.3 Biographies (Experiment 2)

The sky is their/her limit: New astronaut gets ready to travel to the ISS

Ashley Green was born in Chicago in 1990, into a family that always encouraged them/her in their/her interest in science. As a child, Ashley was fascinated by astronomy and stories of space travel. They/She learned to fly a plane before learning how to drive and spent as much time as possible in the sky. Ashley got their/her bachelor's degree in mechanical engineering at the California Institute of Technology and their/her PhD in aerospace engineering at Stanford University.

In 2016, Ashley was one of over 18,000 candidates applying to the astronaut training program at NASA, and one of only 10 candidates chosen. Training to be an astronaut takes two years, which Ashley spent flying training jets, learning medical and survival skills, and learning Russian. To get used to zero gravity, Ashley and other recruits flew in a plane that creates around 20 seconds of weightlessness each time it climbs and dives. Because so many people get sick experiencing zero gravity for the first time, astronauts jokingly call this the "vomit comet" or the "weightless wonder." To practice spacewalks, Ashley spent hours working underwater in a massive pool. They/she say/says that astronaut training was exhausting, but it was exciting to be working with such a passionate and talented team.

In two months, Ashley will be lifting off for the first time, headed to the ISS for a four-month stay. They/She will be working with a team of 5 other astronauts from 3 countries. Ashley's primary job will be upgrading and testing robotic tools used to conduct experiments and do maintenance on the ship. Life on the ISS is busy and isolating, but Ashley says they/she are/is excited to be one the few people to see the earth from orbit.

(underlining added)

An animal-lover's journey to keeping their/his own beehives

James Wright was born in Ohio in 1985. They/he loved animals from an early age, since they/he grew up with their/his family's dogs, cats, chickens, and horses. James began volunteering at the local animal shelter in middle school and caring for foster dogs in high school.

Their/His love of the animal world led James to study biology at The Ohio State University. During this time, they/he also became interested in botany and wanted to learn more about the relationships between plants and animals that exist together in the same environment. A friend recommended a local organization of beekeepers, and James found the perfect combination of their/his interests.

Today, James owns their/his own bee farm, where they/he raise/raises hives to sell to other beekeepers and produce honey. They/He are/is most excited about the free classes they/he run/runs for the community about creating gardens friendly to pollinating insects, caring for your own beehive, and identifying local plants and animals. James hopes to inspire people to learn more about the environments around them and work towards a sustainable future.

Outside of their/his farm business, James is kept busy with their/his 5- and 2-year-old daughters, 3 dogs, and an ever-changing number of cats. They/He also enjoy/enjoys hiking, photography, woodworking, and trying new recipes that use fresh honey.

(underlining added)

B.4 Singular *They* and TGD Attitudes Questions (Experiments 3 & 4)

B.4.1 Naturalness Ratings

Likert scale with 1 = "very unnatural" and 7 = "very natural". Subset of items from Conrod (2019b)*.*

List	Reference Type	Sentence
1	masculine name	Ivan is very generous. They usually have a spare dollar to give to a good cause.
1	feminine name	Reba is very clumsy. They can't try a new sport without getting injured.
1	gender-neutral name	Hayden is very funny. They can make a pun out of almost any word.
1	quantifier	Concerts are very popular. Every music fan tries to buy their ticket early.
1	quantifier	Dogs are very cute. Each dog owner tries to take their dog's picture constantly.
1	generic	The ideal barista is very efficient. They will always make drinks carefully and quickly.
2	masculine name	Mark is very generous. They usually have a spare dollar to give to a good cause.
2	feminine name	Sarah is very clumsy. They can't try a new sport without getting injured.
2	gender-neutral name	Jordan is very funny. They can make a pun out of almost any word.
2	quantifier	Concerts are very popular. Every music fan tries to buy their ticket early.
2	quantifier	Dogs are very cute. Each dog owner tries to take their dog's picture constantly.
2	generic	The ideal barista is very attentive. They will always make drinks carefully and quickly.

B.4.2 Prior Familiarity

Some people use they/them pronouns instead of he/him or she/her pronouns. An example of this is “Alex raised their hand.” How familiar are you with this?
(select all that apply)

- ☐ I use they/them pronouns for myself.
- ☐ I am close to someone who uses they/them pronouns.
- ☐ I have met someone who uses they/them pronouns, but am not close to them.
- ☐ I have heard about people using they/them pronouns, but have not met someone who does.
- ☐ I had not heard about people using they/them pronouns before this study.

[Experiment 3 only] Sometimes people say what kinds of gendered language they want other people to use for them. One way of doing this is saying what pronouns you use when you introduce yourself, like “I’m Emily and I use she/her.” How familiar are you with this?

People tend to include their pronouns when introducing themselves in ___ of the groups I am in.

- ○ All
- ○ Most
- ○ Some
- ○ A few
- ○ None

I include my pronouns when introducing myself:

- ○ Always
- ○ Usually
- ○ Sometimes
- ○ Rarely
- ○ Never, because I prefer not to
- ○ Never, because I had not heard of this before

[Experiment 3 only] Another way of communicating what kind of gendered language you want people to use for you is by including pronouns in places where you write your name, like on nametags, display names, social media bios, and email signatures. How familiar are you with this?

People tend to include their pronouns in places like nametags, display names, bios, and signatures in ___ of the groups I am in.

- ○ All
- ○ Most
- ○ Some
- ○ A few
- ○ None

I include my pronouns in places like nametags, display names, bios, and signatures:

- ○ Always / whenever possible
- ○ Usually
- ○ Sometimes
- ○ Rarely
- ○ Never, because I prefer not to
- ○ Never, because I had not heard of this before

B.4.3 Transphobia Scale

Likert scale with 1 = "completely disagree" and 7 = "completely agree". Items from Nagoshi et al. (2008)*.*

1. I don't like it when someone is flirting with me, and I can't tell if they are a man or a woman.
2. I think there is something wrong with a person who says that they are neither a man nor a woman.
3. I would be upset, if someone I'd known a long time revealed to me that they used to be another gender.
4. I avoid people on the street whose gender is unclear to me.
5. When I meet someone, it is important for me to be able to identify them as a man or a woman.
6. I believe that the male/female dichotomy is natural.
7. I am uncomfortable around people who don't conform to traditional gender roles, e.g., aggressive women or emotional men.
8. I believe that a person can never change their gender.
9. A person's genitalia define what gender they are, e.g., a penis defines a person as being a man, a vagina defines a person as being a woman.

B.5 Character Images (Experiments 3 & 4)

B.5.1 Norming Study

Image Norming: Images from The Gender Spectrum Collection *included.*

Code	Image Title in Gender Spectrum Collection
gs01	A group of friends of varying genders gossiping (second from left)
gs02	A group of friends of varying genders playing with popcorn at a party (left)
gs03	A transfeminine non-binary person and transmasculine gender-nonconforming person looking at a phone and laughing (left)
gs04	A transgender woman sitting on a therapist's couch and smiling
gs05	A transgender woman sitting with an altar
gs06	A transmasculine gender-nonconforming person sitting on a bed
gs07	A transmasculine person in a winter coat on the sidewalk
gs08	A genderqueer person sitting on a therapist's couch listening
gs09	Two non-binary friends playing video games (left)
gs10	A non-binary femme student talking to a friend in front of their locker (right)
gs11	A transmasculine doctor in front of his computer
gs12	Two non-binary students looking at each other in a school hallway photoshopped (left)

Image Norming: Sentence completion prompts. Combinations of images and prompts varied across 3 lists.

Norming Study: Sentence Completion Prompts
This person chose a playlist for the drive. After turning up the volume...
This person collected the trash. After taking it out...
This person got home from work and cooked dinner. After washing the dishes...
This person got out of the pool. After drying off...
This person made a mug of tea. Before it cooled...
This person ordered pizza. Before it arrived...
This person painted the room. While the walls dried...
This person played with a cat. After watching it chase the laser pointer...
This person received a text. After reading it...
This person went on a run at the local park. Before finishing...
This person went to see a show. During the intermission...
This person woke up early to walk the dog. After making coffee...

Table B1: Image Norming: Results. Counts and proportions of they/them, he/him, she/her, and no pronoun responses for each image in norming study.

Norming Study: Pronouns Produced

	they/them	he/him	she/her	none
Images Included				
gs04	19	0	6	5
gs06	20	0	5	5
gs12	24	0	4	2
gs09	24	4	0	2
gs08	22	5	0	3
gs11	21	7	0	2
Images Not Included				
gs03	20	0	7	3
gs05	20	0	6	4
gs02	19	1	6	4
gs07	23	1	2	4
gs10	23	1	3	3
gs01	20	6	1	3
Totals				
Count	255	25	40	40
Proportion	0.71	0.07	0.11	0.11

B.5.2 Images Included

Figure B1. Character Images. Sets of character images used Experiments 3 & 4 and sibling images used in Experiment 3. ST indicates images shown with she/her and they/them pronouns, and HT indicates images shown with he/him and they/them pronouns.

The images for the characters in Experiments 3 and 4 were selected from *The Gender Spectrum Collection*, a stock photo library intended to provide a diverse range of images of transgender and nonbinary people (Drucker, 2019). They are available for use under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International license (CC BY-NC-ND 4.0). Each image was cropped to show one person from the shoulders up with a white background. The titles of the original images are listed below, and the edited images are available upon request.

Character Images From The Gender Spectrum Collection

Character Code	Image Title
ST1	A transgender woman sitting on a therapist's couch and smiling
ST2	Two non-binary students looking at each other in a school hallway photoshopped (left)
ST3	A transmasculine gender-nonconforming person sitting on a bed
HT1	Two non-binary friends playing video games (left)
HT2	A genderqueer person sitting on a therapist's couch listening
HT3	A transmasculine doctor in front of his computer
Information	
Photos	https://web.archive.org/web/20210209092233/https://genderphotos.vice.com/
Usage Guidelines	https://genderspectrum.vice.com/about

The images for the siblings in Experiment 3 were selected from Pexels, a collection of free-use stock photos. Each image was cropped to show one person from the shoulders up with a white background. The creators and titles of the original images are listed below, and the edited images are available upon request.

Sibling Images

Character Code	Creator	Image Title & Source
ST1 Brother	Polina Tankilevitch	A boy wearing a brown hoodie jacket
ST1 Sister	Amina Filkins	Calm African American girl in stylish clothes standing against blue background
ST2 Brother	Kidrays JR	A cute boy with a beautiful smile
ST2 Sister	Kindel Media	A young girl sitting on the sand while holding her toy
ST3 Brother	Mikhail Nilov	Close up shot of a redheaded handsome boy in white long sleeves smiling while posing
ST3 Sister	Andrea Piacquadio	Happy cute small girl in glasses sitting on table in light living room
HT1 Brother	Mikhail Nilov	Boy in white crew neck shirt holding a book
HT1 Sister	Arina Krasnikova	Close up photo of happy cute girl
HT2 Brother	Pavel Danilyuk	A boy in polo shirt holding colored pens
HT2 Sister	Yi Lu	Little schoolgirl sitting behind desk in glasses smiling
HT3 Brother	Victoria Borodinova	Boy wearing white and green t-shirt and green pants
HT3 Sister	Skyler Ewing	Glad girl in pink top
Information		
License	https://www.pexels.com/license/	

The object images for Experiment 3 were selected from stock photos available for use under a Creative Commons Attribution-NonCommercial 4.0 International license (CC BY-NC 4.0).

Object Images

Item	Creator	Image Title & Source
Apple	Scott Ward	Mcintosh Crisp Apple Pie
Avocado	Lydia Simmons	Avocado Half
Bacon	Lydia Simmons	Bacon
Banana	Brett Croft	Banana
Bread	Lydia Simmons	Bread
Broccoli	Scott Ward	Broccoli
Carrot	Alexis Bailey	Carrot
Cherries	Scott Ward	Red Cherry
Chocolate	Hannah Hill	Bar Milk Candy Chocolate
Cookie	Hannah Hill	Cookie
Corn	Hannah Hill	Corn
Egg	Hannah Hill	Egg
Fork	Brett Croft	Fork Spoon and Knife
Grapes	Hannah Hill	Isabella Grape Concord Purple Vine
Hot Dog	Brett Croft	Hot Dog
Ice Cream	Lydia Simmons	Ice Cream
Juice	Scott Ward	Juice
Kiwi	Hannah Hill	Green Cutted Kiwi
Knife	Brett Croft	Fork Spoon and Knife
Lemon	Lydia Simmons	Lemon
Onion	Hannah Hill	Onion
Orange	Alexis Bailey	Orange
Peach	Lydia Simmons	Peach
Pear	Scott Ward	Pear
Pepper	Alexis Bailey	Capsicum Pepper Green Bell
Pineapple	Hannah Hill	Pineapple
Pizza	Lydia Simmons	Pizza
Plate	Lydia Simmons	Red Plate
Popcorn	Alexis Bailey	Popcorn
Potato	Hannah Hill	Potato
Pumpkin	Alexis Bailey	Real Pumpkin
Sandwich	Lydia Simmons	Sandwich
Spoon	Brett Croft	Fork Spoon and Knife
Strawberry	Alexis Bailey	Strawberry
Tomato	Alexis Bailey	Tomato
Watermelon	Lydia Simmons	Watermelon
Information		
License	https://freepngimg.com/about#license	

B.6 Stories (Experiment 4)

S01	Alex & Sam / Sam & Jordan	[Name] is painting a portrait of [Name]	and some paint is spilling on the floor	and the painting looks amazing
S02	Alex & Sam / Sam & Jordan	[Name] is playing the cello for [Name]	as the telephone is ringing	but it doesn't look like anyone is going to answer the phone
S03	Alex & Sam / Sam & Jordan	[Name] is making a salad for [Name]	while a pot is boiling on the stove	and it looks like dinner is almost ready
S04	Alex & Sam / Sam & Jordan	[Name] is walking over to [Name]	and a broken radio is lying on the ground	but it doesn't look like the radio can be fixed
S05	Alex & Sam / Sam & Jordan	[Name] is picking up [Name] from school	as the bell is ringing	and the bell is very loud
S06	Alex & Jaime / Jordan & Taylor	[Name] is bringing some mail to [Name]	as a violent storm is beginning	and it looks like a bad day to go outside
S07	Alex & Jaime / Jordan & Taylor	[Name] is doing a dance for [Name]	while the music is blasting out of a record player	and the chandelier is shaking from all the noise
S08	Alex & Jaime / Jordan & Taylor	[Name] is taking a picture of [Name]	next to the Eiffel Tower	and the tower is lit up at night
S09	Alex & Jaime / Jordan & Taylor	[Name] is playing golf with [Name]	as the sun is shining overhead	and the grass is very green
S10	Alex & Jaime / Jordan & Taylor	[Name] is floating over to [Name]	while an airplane is flying overhead	and it looks like everyone needs more sunscreen
S11	Alex & Jordan	[Name] is singing to [Name]	and there's a big disco ball overhead	and all the colored lights look fun

S12	Alex & Jordan	[Name] is showing a happy birthday sign to [Name]	while some balloons are floating up to the ceiling	and it looks like the party is starting
S13	Alex & Jordan	[Name] is surfing at the beach with [Name]	and a crab is scuttling by	and the waves keep getting bigger
S14	Alex & Jordan	[Name] is racing toward [Name]	while a house is burning down	and it sounds like the fire truck is coming
S15	Alex & Jordan	[Name] is looking at [Name]	as the train whistle is blowing loudly	but the whistle is too loud to hear anything
S16	Alex & Taylor	[Name] is skateboarding toward [Name]	as a car is speeding down the road	and it looks like there might be an accident
S17	Alex & Taylor	[Name] is singing a song to [Name]	as the moon is high in the sky	and it sounds like the song is almost over
S18	Alex & Taylor	[Name] is rollerblading towards [Name]	while several birds fly overhead	but it looks very easy to fall down
S19	Alex & Taylor	[Name] is trying to teach some algebra to [Name]	as the clock reads three	and it appears that the lecture is about to end
S20	Alex & Taylor	[Name] is shouting at [Name]	while a shooting star streaks across the sky	but it looks like an argument has started
S21	Alex & Casey	[Name] is carrying a cake to [Name]	as the refrigerator starts to hum	and it looks like the cake is fresh out of the oven
S22	Alex & Casey	[Name] is filming a video for [Name]	as clouds start to cover up the sun	but it looks like it might start to rain soon
S23	Alex & Casey	[Name] is talking to [Name]	while a spotlight is shining overhead	and it looks like rehearsal just started
S24	Alex & Casey	[Name] is skiing with [Name]	and snow is falling from the sky	and it looks like everyone might be cold
S25	Alex & Casey	[Name] is dancing with [Name]	while a pinata is hanging overhead	and everyone at the party is having fun

S26	Sam & Taylor	[Name] is bringing some balloons to [Name]	while bright flowers are sitting in the window	and it must be someone's birthday
S27	Sam & Taylor	[Name] is walking with [Name]	and the clock is about to ring 3 o'clock	and it seems like class will be starting soon
S28	Sam & Taylor	[Name] is flying on a broom above [Name]	while a bat flutters through the air	and it looks like it is Halloween
S29	Sam & Taylor	[Name] is playing basketball with [Name]	as the scoreboard shows the time winding down	and the game appears to be one-sided
S30	Sam & Taylor	[Name] is playing the violin for [Name]	while the sun is shining brightly	and it looks like the song is being played well
S31	Sam & Casey	[Name] is practicing soccer with [Name]	as a bird flies overhead	and it looks like the ball just missed the net
S32	Sam & Casey	[Name] is cheering on [Name]	as the wind is blowing some leaves	and it looks like a race is about to start
S33	Sam & Casey	[Name] is calling [Name]	while a cat is walking by	and it looks like the cat is very friendly
S34	Sam & Casey	[Name] is buying a gift for [Name]	but it is almost lunchtime	and there might not be a chance to buy the gift today
S35	Sam & Casey	[Name] is swimming towards [Name]	and a goose is swimming nearby	and it is a beautiful sunny day
S36	Jordan & Casey	[Name] is scuba diving with [Name]	and a big fish just came out of nowhere	and it looks like there's a lot more fish about to come by
S37	Jordan & Casey	[Name] is looking for [Name]	and it looks like it's about to rain	and it's past time to get inside
S38	Jordan & Casey	[Name] is exercising with [Name]	and the time is going by really fast	but the gym is about to close
S39	Jordan & Casey	[Name] is taking a selfie with [Name]	and the room is kind of dark	but the picture is coming out ok

S40	Jordan & Casey	[Name] is bringing a coffee to [Name]	since it's very early in the morning	and it looks like it's almost 6am
S41	Jordan & Jaime	[Name] is reading to [Name]	and the story is almost over	and it sounds like the story is a new favorite
S42	Jordan & Jaime	[Name] is sending a text to [Name]	as the battery is about to die	but it looks like the text still sent
S43	Jordan & Jaime	[Name] is getting a chair for [Name]	because the wait will be a while	but it looks like the wait will be worth it
S44	Jordan & Jaime	[Name] is showing a picture to [Name]	while the dog is coming over	and the dog is just as cute as in the picture
S45	Jordan & Jaime	[Name] is explaining the game to [Name]	since it's kind of complicated	and it looks like the game will be fun
S46	Sam & Jaime	[Name] is giving flowers to [Name]	after the show is over	and all the performers were talented
S47	Sam & Jaime	[Name] is driving away from [Name]	while the sun is starting to set	and it looks like it's a long drive home
S48	Sam & Jaime	[Name] is playing a video game with [Name]	and it's starting to get late	but the final level hasn't been beaten yet
S49	Sam & Jaime	[Name] is whispering to [Name]	as the movie is starting	and everyone is quieting down
S50	Sam & Jaime	[Name] is throwing the ball to [Name]	while the dog is getting excited	and the dog caught the ball instead
S51	Taylor & Casey / Casey & Jaime	[Name] is starting a video call with [Name]	but the connection is kind of slow	and it's great to see each other
S52	Taylor & Casey / Casey & Jaime	[Name] is passing a blanket to [Name]	because the room is cold	and everyone is comfortable now

S53	Taylor & Casey / Casey & Jaime	[Name] is making a smoothie for [Name]	before it's time to leave	and it's hard to hear above the blender
S54	Taylor & Casey / Casey & Jaime	[Name] is opening the door for [Name]	as the keys rattle in the lock	and no one else is in the house
S55	Taylor & Casey / Casey & Jaime	[Name] is holding a spot for [Name]	because the line is moving slowly	and the day keeps getting hotter
S56	Taylor & Jaime	[Name] is carrying a box for [Name]	since it's very heavy	but there's not much left to move
S57	Taylor & Jaime	[Name] is giving directions to [Name]	and the traffic is starting to clear	and it looks like the rest of the trip will be easy
S58	Taylor & Jaime	[Name] is telling a joke to [Name]	while the party is getting started	and everyone is laughing at how funny it is
S59	Taylor & Jaime	[Name] is taking notes for [Name]	under the fluorescent lights	and it looks like the meeting is almost finished
S60	Taylor & Jaime	[Name] is hiking along the trail with [Name]	and the leaves are starting to turn	and the top of the mountain isn't much further