Gender competition in the production of nonbinary 'they'

Jennifer E. Arnold, Ranjani Venkatesh, and Zachary A. Vig

Department of Psychology and Neuroscience, The University of North Carolina Chapel Hill

Word count: 13796

Correspondence concerning this article should be sent to Jennifer Arnold.

Email: jarnold@unc.edu

Abstract

Two experiments test how college students use nonbinary *they* to refer to a single and specific person whose pronouns are *they/them*, e.g., "Alex played basketball on the neighborhood court. At one point they made a basket," compared to matched stories about characters with binary (she/her or he/him) pronouns. Experiment 1 shows that for both types of pronouns, people use pronouns more in a one-person than a two-person context. In both experiments, people produce nonbinary *they* at least as frequently as binary pronouns, suggesting that any difficulty does not result in pronoun avoidance in spoken language, even though it does in written language (Arnold et al. 2022). Nevertheless, there is evidence that nonbinary *they* is somewhat difficult in that people made gender errors on about 9% of trials, and they used a more acoustically prominent and disfluent-sounding pronunciation for nonbinary than binary pronouns. However, exposure to *they* in the context of the experiment had no effect on frequency, accuracy, or pronunciation of pronouns. This provides the first evidence of how nonbinary *they* is used in a naturalistic storytelling context and shows that while it poses some minor difficulties, it can be used successfully in a supportive context.

Key words: pronoun, nonbinary they, production

1. Introduction

The pronoun *they* is currently undergoing a change in how it is used. In the last decade, there is a growing awareness that some individuals use *they* as their pronoun of reference, for example "Demi announced… *they* are nonbinary" [italics added] (Bate, 2021, about Demi Lovato). This change is supported by the trend to talk about pronouns, e.g. "my pronouns are they/them." In many cases *they* is used by individuals who identify as gender nonbinary or gender queer, so for convenience we call it "nonbinary *they*".

The entrance of nonbinary *they* has the potential to dramatically change both the English language processing system and mainstream concepts of what gender is (and conversely, changing concepts of gender may influence the pronominal system). Pronouns are highly frequent words, so they are bound to be used in many situations where a *they/them* user is mentioned. But notably, this usage is at odds with the grammars of some speakers (Bjorkman, 2017; Konnelly & Cowper, 2020). So how does the system adapt to nonbinary *they*? Here we address this question by examining the behavior of speakers engaged in a story-telling task where they refer to characters who use *she/her, he/him*, and *they/them* pronouns.

Our goal is to test the ways that the production of nonbinary *they* is similar to or different from the production of binary *he* and *she* for a sample of young adult speakers at this point in time. It is notable that at the time of running these experiments in 2021-2022, nonbinary singular *they* is still relatively new. Some people argue against it either because they view it as ungrammatical or are ideologically opposed to nontraditional genders (Ben, 2019; for discussion see Conrod, 2020). Yet many published views and institutional policies work in favor of both inclusive language in general and singular *they* in particular.[1] Our study critically focuses on a

[1] For example, UNC's office of Diversity and Inclusion states that "UNC-Chapel Hill strives to ensure gender equity across all platforms, including hiring practices, lactation/family support, and gender-inclusive language," and that

university community (UNC Chapel Hill) where inclusive language is publicly valued, and the nonbinary *they* form is familiar to students. Yet even in this context, nonbinary *they* is still relatively new and low frequency. This low frequency could potentially disrupt the process of selecting pronouns when appropriate to the context, lead to errors, or result in disfluent production.

It is important to understand how *they* is used in naturalistic language production because research shows that gender diverse individuals experience a high risk for mental health problems (Gross et al., 2022) and misgendering causes distress (McLemore, 2015), but proper pronoun use significantly reduces this risk (Sevelius et al., 2020). The use of *they* in appropriate contexts may be especially socially salient because it clearly stands out in contrast to binary pronouns. Thus, an increase in *they* use has the potential to have a positive impact on public health. Yet even people who wish to use *they* respectfully may find it difficult. In this study we aim to understand the extent to which production patterns of *they* differ from binary *she* and *he*, and in what ways. From a theoretical perspective it helps us understand the process of adapting to a new and societally-relevant form. From a practical perspective this work has the potential to guide efforts to improve fluency with nonbinary *they*.

Indeed, evidence suggests that the comprehension of singular *they* can lead to processing disruptions. In event-related potential studies, singular *they* or *themselves* can elicit a P600, which is associated with syntactic anomalies (Leventhal et al., 2020; Prasad & Morris, 2020), and *they* is easier to understand with a plural than a singular interpretation (Sanford & Filik,

---

"Asking and correctly using someone's pronouns is one of the most basic ways to show respect for their individuality and gender identity." (as of April 7, 2023; https://diversity.unc.edu/gender-equality/). As another example, the University of Minnesota officially states that "University members and units are expected to use the names, gender identities, and pronouns specified to them by other University members, except as legally required. University members and units are also expected to use other gendered personal references, if any, that are consistent with the gender identities and pronouns specified by University members." (https://policy.umn.edu/operations/genderequity).

2007). However, this disruption is limited to cases where the referent has an assumed gender. People read singular *they* relatively quickly if the referent is generic (e.g., *anyone; a runner*) vs. specific (*My nurse*) or named (*Chloe*; Ackerman, 2018; Foertsch & Gernsbacher, 1997), or if the referent has no expected gender (*the cyclist)* vs. has a stereotypical gender (*the mechanic;* Doherty & Conklin, 2017). Moulton et al. (2022) found that regardless of gender, people found singular *they* more natural and easier to process when preceded by a quantified antecedent that signals distributivity (e.g., *Each cyclist* vs. *All the cyclists).* This line of work suggests that singular *they* is only hard to understand when readers assume the antecedent has a binary gender.

Current theories suggest that individuals vary in their acceptance of singular *they* with different sorts of antecedents. Konnelly and Cowper (2020) build on Bjorkman's (2017) account to suggest that English speakers fall into three categories, which they define in terms of the grammatical algorithms governing the selection of one gender pronoun over another. Individuals in their Stage 1 allow *they* for reference to quantified entities (e.g., "Every student must turn in *their* homework"). This usage has been established for centuries, and is attested in the writings of Shakespeare, Austen, and others (Baron, 2020; Nunberg, 2016; McWhorter, 2018). Their Stage 2 allows for *they* to refer to entities introduced by ungendered nouns, e.g. "The teacher said they needed a break," but assumes that some nouns are gendered and disallows examples like "My mother said they were tired." On their Stage 3, *they* may refer to any singular person, regardless of gender. Camilliere et al. (2021) collected data on the acceptability of *they* in different sentence contexts and used it to show that respondents indeed clustered into three groups, which they termed "non-innovators", "innovators", and "super-innovators." Only their super-innovator group accepted *they* used to refer to any singular animate referent, including gendered descriptions (*my sister…they*) or names (*Sophia…they*). It is also clear that acceptance of

singular *they* varies systematically with demographic factors. The more innovative users tend to be younger (Camilliere et al., 2021; Conrod, 2019) and more familiar with nonbinary people (Ackerman, 2018; Bradley et. al., 2020).

Notably this line of work examines the changing use of *they* through the lens of the grammar, with the idea that individuals may move from one grammar to another as a function of new input. This approach treats linguistic knowledge as categorical – a speaker either does, or does not, claim to accept a particular usage. Yet these stated patterns of acceptance may differ from the way people actually speak. People may label uses of *they* for the antecedent "the teacher" as ungrammatical if they believe they were taught that this is incorrect, but nevertheless use singular *they* in spoken language. Alternatively, people may accept "Sophia" as an antecedent for *they* because they value the inclusivity of gender-neutral *they* and still struggle to use it.

Thus, an unanswered question is how speakers actually use *they* in discourse. Do they produce singular *they* when appropriate and in a similar way to *she* and *he*? It is well established that people use pronouns in specific discourse situations, for example when the referent has been recently mentioned or is in a prominent linguistic position (e.g., Ariel, 1990; Arnold, 1998; Arnold & Zerkle, 2019; Chafe, 1976; Gundel et al., 1993). For example, pronouns are frequently used when referring to the subject of the previous clause, but more often when there is a single person in the story than two (e.g., in *Mickey went for a walk…. He…* vs. *Mickey went for a walk with Daisy…. He…* . ; Arnold & Griffin, 2007). Once nonbinary *they* is fully integrated into the language, we would expect the production of nonbinary *they* (vs. names or descriptions) to occur at the same rate as binary pronouns and in similar discourse conditions.[2] On the other hand, for

[2] For example, discourse prominence should have similar effects on both binary and nonbinary pronouns, and both should be impacted by competition with other potential referents in the context, although what counts as a

many people, nonbinary *they* is low frequency and unpracticed, which may lead to differences in its usage.

One possibility is that speakers may avoid using nonbinary *they* and instead use names, even when a pronoun would be appropriate. This is precisely the pattern observed for written language in a text analysis. Arnold et. al. (2022) examined writers' choices between pronouns and more explicit expressions (names, descriptions), comparing the production of nonbinary *they* and binary *he/she* while controlling for discourse context (given vs. new). Their analysis focused on 27 published articles about nonbinary individuals, which represent real-life cases where the writer knows about nonbinary *they* and wishes to use it. These were compared to binary *he/she* references from the same authors. For each binary and nonbinary target character, they analyzed the first singular and non-possessive reference occurring in each sentence, excluding the first one in the article (which is always nonpronominal). Both nonbinary and binary pronouns were used more often when the referent was "Given" (mentioned in the previous sentence) than when it was "New" (not mentioned in the previous sentence), exhibiting the well-known tendency to use pronouns more for given than new information (Ariel, 1990; Chafe, 1976; Gundel et al., 1993). This suggests that in general people use the same constraints for selecting binary and nonbinary pronouns. But critically, nonbinary pronouns were produced less often than binary pronouns. This pattern did not result from differences in the ambiguity of binary and nonbinary pronouns; the same effect was observed for tokens where there were no competing referents in the previous sentence. Importantly, using a name is not socially offensive; it is a perfectly acceptable term of

competitor for "she" is the presence of another female referent, while what counts as a competitor for "they" would include the presence of another they-user and/or a plural referential group that could be referred to with "they".

reference. However, this pattern reflects a different decision-making process for *they* vs. *she/he* pronouns.

The authors considered two possible explanations for this finding. First, all references involve a selection process, and the pronoun may have been only weakly activated for nonbinary antecedents because of its low frequency, leading to a greater likelihood of selecting the more explicit name or description. Second, writers may have suppressed their use of nonbinary *they* out of concern that some readers may not be as familiar with it and find it difficult to understand.

Thus, Arnold et al.'s (2022) text analysis suggests that producing nonbinary *they* may be somewhat harder than producing binary pronouns. Does the same effect occur in spoken language? To our knowledge there is no evidence in the literature to answer this question.

Here we present two storytelling experiments that probed the use of both binary and nonbinary pronouns. We test two questions. First, do people make similar decisions about when and how frequently to use binary and nonbinary pronouns? To assess this, we examined whether people produced a pronoun or a name for the target character. One part of this question was whether the discourse context affects binary and nonbinary pronoun use similarly. We predict it does, given the findings from written production (Arnold et al., 2022). It is well known that binary pronouns are more likely to be used in the context of a single character than two characters, especially when the two characters have the same gender (Arnold & Griffin, 2007). Here we examine whether the number of characters guides both binary and nonbinary pronouns in the same way. Another part of this question is whether people produce binary and nonbinary pronouns at an equal rate, after controlling for discourse context. If spoken language is like written language, people may over-produce names for nonbinary referents. Alternatively, spoken language is different from written language in several ways: speakers have less time to evaluate

their productions and edit them, whereas writers can revise as much as needed before publication. In addition, face-to-face conversation often takes place with a specific addressee, whereas writing is available to a broader audience. Both differences may impact pronoun production.

Our second question is whether there is evidence that nonbinary pronouns are harder to produce than binary pronouns. To assess this, we analyze two things. First, we use pronunciation as an indicator of fluency, measured through perceptual ratings. Second, we examine gender errors in pronoun choice to assess whether errors are more common for nonbinary than binary pronouns.

Fluently produced pronouns tend to have reduced prominence, especially when produced in a context that highly supports pronouns – such as all the contexts examined here. By contrast, when people are disfluent, they tend to slow down and use more prosodically prominent pronunciations (see Arnold & Watson, 2015; Kahn & Arnold, 2012). Prosodic prominence is frequently analyzed in terms of its relation to linguistic structure, such as whether a word is accent or not (e.g., Ladd, 1986; Cole, Mo & Johnson, 2010), and acoustic variation reflects information status such that given information tends to be more reduced than new information (Halliday, 1967). But prominence is not a pure representation of linguistic structure, and prominent pronunciations also reflect processing load associated with speech planning (Arnold & Watson, 2015; Arnold, Kahn, & Pancani, 2011; Bell et al., 2003; Ferreira & Swets, 2002; Kahn & Arnold, 2012; Watson et al., 2008). In the current experiment, we test production of referential expressions in discourse contexts where the target is always given and informationally salient as the subject of the prior sentence. Thus, any observed variation in prosodic prominence is likely to stem from processing differences. If people are having difficulty selecting and

retrieving nonbinary pronouns, we expect that nonbinary pronouns will be uttered with a more emphatic and prosodically prominent pronunciation than binary pronouns.

An important feature of this study is that it examines pronoun use in a context where nonbinary *they* is pragmatically supported. Linguistic accounts suggest that for the most innovative users, singular *they* is grammatical regardless of the gender identity of referent (Bjorkman, 2017; Konnelly & Cowper, 2020). However, part of knowing a language goes beyond just knowing what is acceptable and includes knowing the pragmatic rules for appropriate usage. Conrod (2020) points out that politeness dictates that speakers use the correct personal pronouns for reference to a person, and misgendering occurs when speakers use someone's dispreferred pronouns. Practically speaking, this means that the likelihood of producing *they* is much higher when the speaker knows that the referent's personal pronouns include *they/them*. Evidence from comprehension shows that people are more likely to interpret *they* with a singular meaning if pronouns have been explicitly introduced, e.g. "Alex uses they/them pronouns" (Arnold et. al., 2021). Our study therefore sets the context for our storytelling task by introducing participants to a fictional cast of characters that includes two people who use *she/her* pronouns, two people who use *he/him* pronouns, and one person whose pronouns are *they/them.*

Table 1

Demographics of Participants

Demographic	Experiment 1		Experiment 2	
	n	%	*n*	%
Gender Identity*				
Male	14	46.2	3	12.5

Female	12	53.9	21	87.5
Race/Ethnicity				
White, Non-Hispanic	19	73.1	10	41.7
White, Hispanic	2	7.7	3	12.5
Asian, Non-Hispanic	2	7.7	4	16.7
Black or African American, Non-Hispanic	1	3.9	1	4.2
More than one race, Non-Hispanic	1	3.9	4	16.7
More than one race, Hispanic	1	3.9	0	0
Asian, Do not wish to report (ethnicity)	0	0	1	4.2
Do not wish to report (race or ethnicity)	0	0	1	4.2
	M	*SD*	*M*	*SD*
Age	19.0	1.0	19.9	3.2
Year in School§	1.7	0.8	1.1	1.4
	M	Mode	*M*	Mode
How many people do you know who identify as nonbinary? (0 to *5 or more*)¤	1.1	0	3.1	2
In how many languages besides English are you conversationally proficient? (0 to *3 or more*) ¤	0.6	0	1.9	0

* All participants reported their sex assigned at birth to be the same as their gender identity.
§ Year in school was coded as 5 for participants who reported year in school as 4+
¤ "5 or more" was coded as 5 and "3 or more" was coded as 3 for calculating the mean

2. Experiment 1

2.1 Methods

2.1.1 Participants

Twenty-nine students from the University of North Carolina at Chapel Hill (see Table 1) participated in exchange for course credit. Three were excluded from analysis: one was a nonnative speaker; one did not give permission to record; and one due to experimenter error. Twenty-six subjects were included in the analysis.

2.1.2 Materials and Design

Using a variation of Arnold & Griffin (2007; see also Zerkle & Arnold, 2019) we presented participants with two-panel cartoons. Participants were instructed to help tell a story

based on the pictures. The beginning of the story was provided in written form below the first panel. They read this prompt out loud and then forwarded to the next panel. This panel also had a short prompt written on screen; they read this prompt and then continued the sentence in their own words based on the picture. The second panel pictured one person doing something interesting; this was the target character.

The stories were all about five people, Liz (she/her); Alex (they/them); Ana (she/her); Will (he/him); and Matt (he/him); see Figure 1. Pictures of these characters, their names and their pronouns were introduced before the main task. Participants were then tested to make sure they remembered the names and pronouns that went with each picture.

There were 24 critical stories in the experiment plus 24 fillers and 4 practice items. There were two within-items manipulations, such that each story appeared in four versions. First, we manipulated whether the story included one or two people. The two-person stories always used the structure "X did something with Y" for the first sentence; the one-person stories said only "X did something". The cartoon panels were identical except the second person was eliminated from the one-person stories. In the two-person stories the image of the second person was the same for both panels, suggesting that this person was not involved in the target action.

Second, we manipulated whether the target character was binary or nonbinary. In the binary condition, the first person mentioned was one of the *he/she* characters, and the second person mentioned (if present) was the other character of the same gender (e.g., Ana and Liz, or Will and Matt). In the nonbinary condition, the first person mentioned was always Alex, and the second person (if present) was one of the other four characters. For all critical stories, the first sentence mentioned the two people in a sentence with the structure "X did something with Y", which we term the "joint action" structure. See Table 2 for an example of the context sentences

and prompts; see Figure 2 for an example of visual stimuli. See supplementary materials on the osf site for stimuli and pictures (see Data Accessibility).

Figure 1

Story Characters: Liz (she); Alex (they); Ana (she); Will (he); Matt (he)

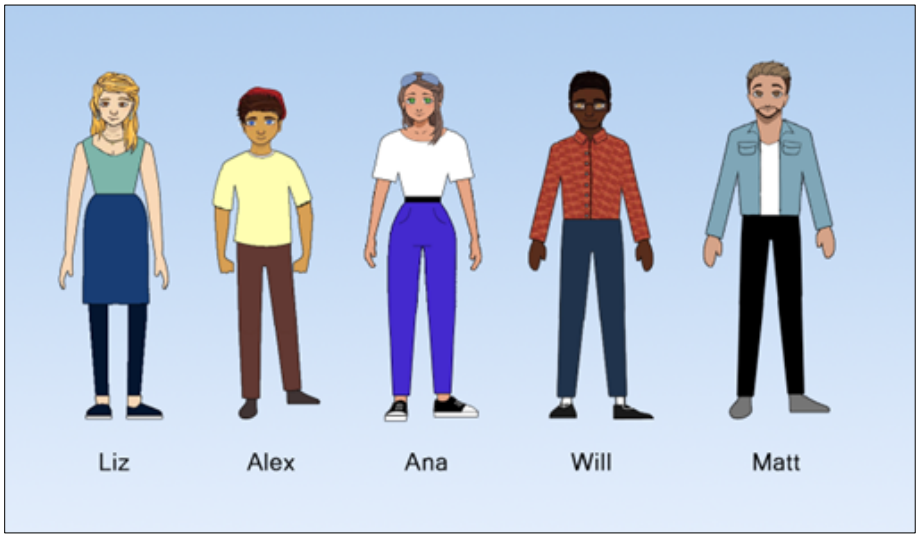


Table 2

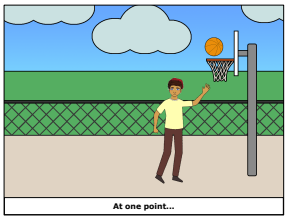
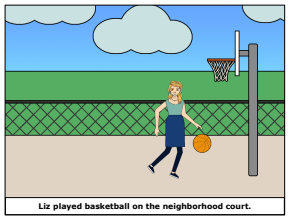
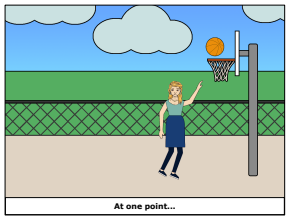
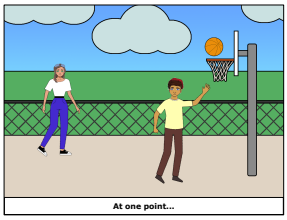
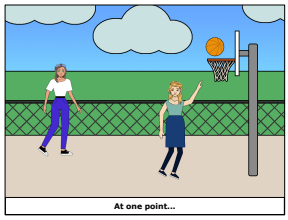
Example Context Sentences and Story Prompts in Each Condition

Condition	Example
Binary / One Person	*Liz played basketball on the neighborhood court. At one point…*
Binary / Two People	*Liz played basketball with Ana on the neighborhood court. At one point…*
Nonbinary / One Person	*Alex played basketball on the neighborhood court. At one point…*
Nonbinary / Two People	*Alex played basketball with Ana on the neighborhood court. At one point…*

The 12 filler stories used contextual structures of different types, for example with two people mentioned in a conjoined subject NP (e.g., *Liz and Ana played a card game all afternoon. Then…*), where the target picture showed both of them doing something together in the next panel, or in a two-person sentence where the second panel illustrated the second-mentioned person (e.g., *Ana listened to Alex recite a poem on stage. After that…*).

Figure 2

Sample Visual Stimuli for "Alex/Liz played basketball {with Ana} on the neighborhood court. At one point…"

Nonbinary Target Condition		Binary Target Condition	
Panel 1	Panel 2	Panel 1	Panel 2
One Character Condition			
			
Two Character Condition			
			

2.2 Procedure

Participants met the experimenter in a one-on-one Zoom session. The experimenter described the task and asked the participant to fill out a Qualtrics survey with a consent form and demographic questions (see supplementary material). The participant was then instructed to turn off their video and change their Zoom name to their participant number so that the recording would be anonymous.

The experimenter introduced the story characters and then tested the participant's memory for the character names and pronouns. If the participant made any mistakes, the experimenter corrected them. The experimenter then began recording, and the participant did four practice items. If the participant used incorrect names or pronouns during the practice items, the experimenter corrected them, e.g. saying "remember, Alex's pronouns are they/them". The

participant was given a chance to ask questions. Then the main task began; after this point the experimenter did not correct any mistakes.

2.3 Results and Discussion

2.3.1 Analytical approach

The binary outcome (pronoun vs. name) was analyzed with a mixed-effects logistic regression using SAS proc glimmix, with a binary distribution and logit link. The quantitative outcome (prosodic rating) was analyzed with a mixed-effects linear regression using SAS proc mixed. Binary predictors were effects-coded 1 vs. -1. All models included random intercepts for subject and item, and maximal slopes as appropriate.

2.3.2 Analysis #1: Pronoun or Name?

For the primary analysis, two coders transcribed participant responses for all practice and critical items. They then identified the target referring expression, which was defined as the referring expression that occurred in the subject position of the response and referred to the target character, and coded whether it was a pronoun (*he, she, they*) or a name.[3]

As shown in Appendix A, 53 trials (8% of the data) were excluded for the following reasons: a) the grammatical subject NP in the response did not refer to the target character, the target event in panel 2 was not described in accordance with the picture, or the participant changed the structure of the response sentence (for example, adding an additional phrase before the target event, e.g. "using Ana's advice uhhh Alex found the artifact they had been looking for"); b) for the target referring expression, the participant used the wrong name or pronoun or

[3] Transcribers were instructed to record each thing said, including part words, disfluencies, and pauses. Most of the responses were transcribed and coded by only one person. To check cross-coder reliability, both coders transcribed/coded participants 1 and 11. Transcriptions were 100% in agreement at the level of word identification. Coding was 98% in agreement.

corrected the expression; c) the first sentence and prompt were not read accurately; minor word changes were allowed but not if it changed the meaning or references, d) the response was not recorded or inaudible; e) the response did not mention the target referent explicitly (e.g., "After that played the piano").

A potential concern in the two-person condition was that participants might describe the actions of both characters together instead of just the target character, even though the second character was always backgrounded and playing a passive role. If a plural reference is produced with the pronoun "they", it could be ambiguous in the nonbinary condition. We therefore examined the binary condition to estimate the degree to which plural responses may have occurred in the nonbinary condition.

Our first question was whether the impact of the discourse context (one vs. two characters) would affect both binary and nonbinary pronouns. As Figure 3 illustrates, participants were much more likely to use both binary and nonbinary pronouns in the one-person context than in a two-person context, replicating the established tendency to use pronouns more frequently in one-person contexts (e.g. Arnold & Griffin, 2007).

Our second question was whether participants would underuse pronouns in the nonbinary context, compared to the binary context. In contrast with findings for published articles (Arnold et al., 2022), we saw no hint of this effect. In fact, we saw the opposite effect, where nonbinary pronouns were somewhat more likely than binary pronouns in the two-character condition.

We examined these patterns with a mixed effects logistic regression. As shown in Table 3, we found a significant effect of the nonbinary predictor, as well as a marginal interaction between nonbinary (vs. binary) and one (vs. two) characters. To probe the marginal interaction, we used estimates to calculate the effect of gender (binary vs. nonbinary) in the one- and two-

person conditions. These revealed that pronoun use was no different for nonbinary and binary conditions when there was only one person in the story (both 78%), but they were significantly different in the two-person condition (20% for nonbinary vs. 9% for binary).

Figure 3

Results from Experiment 1: Rate of Pronoun Use in Each Condition

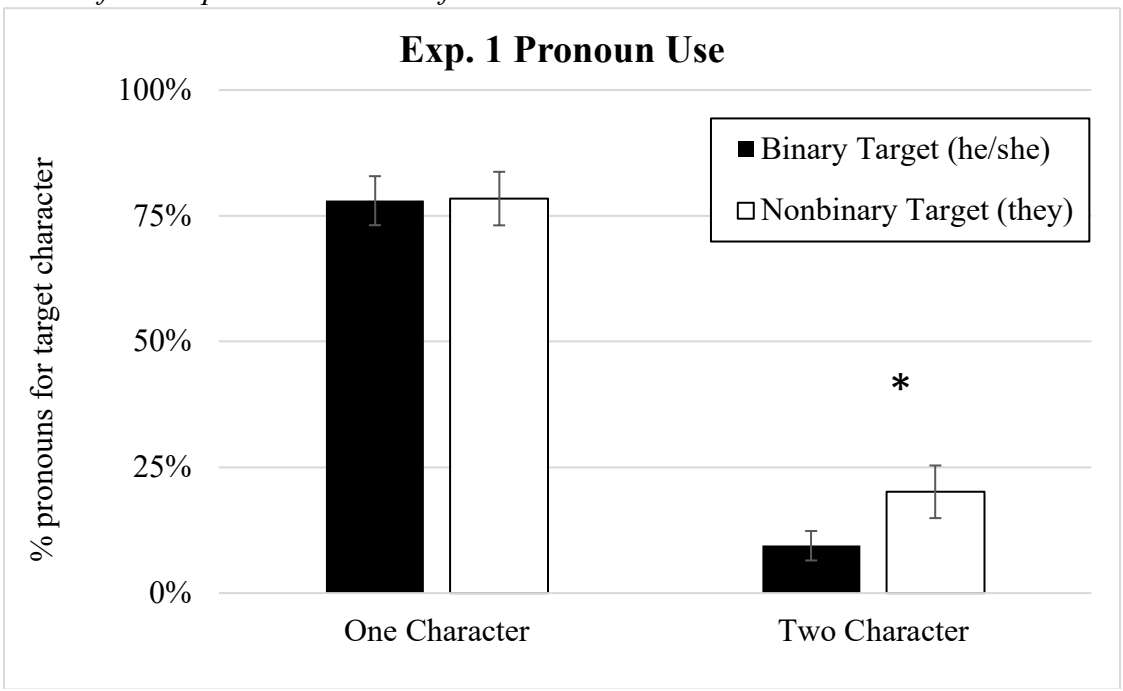


Table 3a

Reference Form Analysis: Inferential Statistics from Experiment 1

Effect	Estimate (Std. Error)	*t* Value	Pr > |*t*|
Intercept	-0.27 (0.33)	-0.82	0.42
Nonbinary vs. Binary Pronoun	0.33 (0.14)	2.37	0.03
One- vs. Two-Character Condition	2.08 (0.17)	11.91	<.0001
Nonbinary * One Character	-0.27 (0.14)	-2.01	0.06

Table 3b

Reference Form Analysis: Estimates of the Nonbinary Effect for One- vs. Two-Character Conditions in Experiment 1

Effect	Estimate (Std. Error)	*t* Value	Pr > |*t*|
One-Character: Nonbinary	0.11 (0.35)	0.31	0.76
Two-Character: Nonbinary	1.2 (0.42)	2.83	0.01

2.3.3 Analysis #2: Prosodic Prominence as a Signal of Fluency

For the trials included in Analysis 1, we coded the perceived prosodic prominence of the name or pronoun. An additional 15 items were excluded because the audio was too poor to identify prosodic prominence, the name was repeated, or the pronoun was corrected (e.g., "them they"). Four raters listened to the recordings of the stories that were included in analysis 1. They coded the perceived prominence of the critical name or pronoun on a scale of 1 – 3 plus half points, resulting in a six-point scale (1, 1.5, 2, 2.5, 3, and 3.5; see Appendix B for coding details). All four codings for each response were averaged for the final analysis.

Our primary analysis compared the pronouns and names produced in the four conditions. Results (Table 4; Figure 4 left panel) revealed that names were perceived as more prominent than pronouns. In addition, nonbinary pronouns were perceived as more prominent than binary pronouns, but there was no difference between binary and nonbinary names. This pattern emerged in our model as an interaction between nonbinary/binary and pronoun/name (see Table 5a). We probed the interaction with estimates and found that pronouns were significantly more prominent in the nonbinary than binary condition, but there was no difference between conditions for names (see Table 5b).

Table 4

Average Prosodic Prominence Ratings by Condition for Experiment 1

Condition		Pronoun	Name
Binary Target (he/she)	One Char	1.677	1.851
	Two Chars	1.596	2.118
Nonbinary Target (they)	One Char	1.836	1.971

	Two Chars	1.784	2.031

In a secondary and post-hoc analysis, we examined whether the pronunciation of "they" differed for singular and plural uses. If the prominence of nonbinary "they" stems from difficulty, it may also be perceived as more prominent than plural uses of "they". This analysis capitalized on the fact that four of the filler items introduced two people in a conjoined NP (e.g., "Liz and Ana"), and pictured them performing an action together in the second panel that was typically described with the plural pronoun "they".[4] The same four coders were asked to code this subset of items over a year after the initial coding (see Appendix B for further details). Numerically we observed greater prominence for nonbinary they (Avg. = 1.90) than plural they. (Avg. = 1.77). This difference between singular and plural they (0.14) is similar to the difference between binary and nonbinary pronouns, averaging across number of characters (0.16). However, this analysis was underpowered since we only had 4 plural fillers, and the difference failed to reach significance (b = 0.05 (SE = 0.04), t = 1.26, p = 0.22).

Figure 4

Average Prominence Ratings for Experiment 1. Ratings for Pronouns and Names in Critical Trials.

[4] The four filler stories and example responses are: 1) *Liz and Ana played a card game all afternoon. Then…they stacked the cards;* 2) *Ana and Will threw a birthday party last night. After that… they washed the dishes;* 3) *Alex and Liz took a canoe trip last weekend. During the trip… they capsized;* 4) *Will and Matt talked all morning at a coffee shop. At one point... they had sandwiches.*

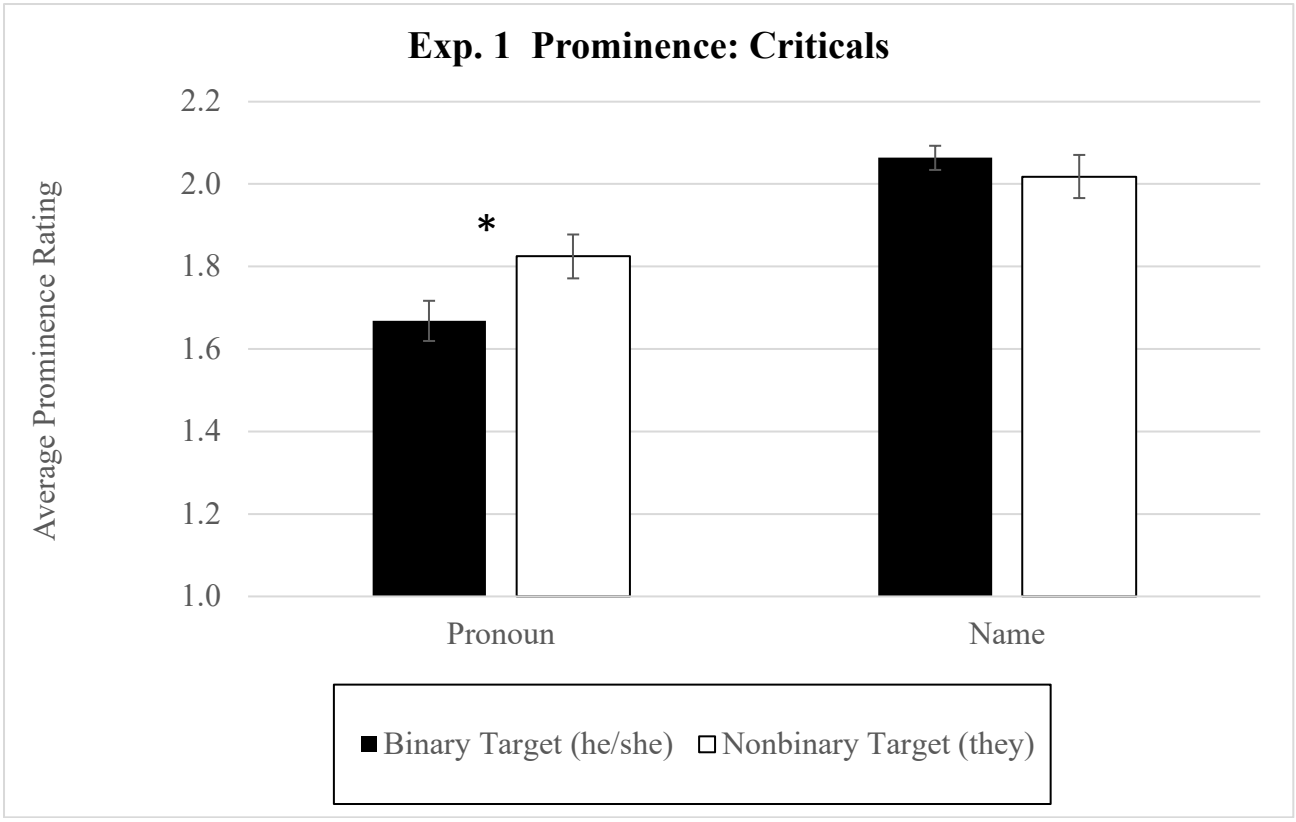

Exp. 1 Prominence: Criticals
Average Prominence Rating
2.2
2.0
1.8
1.6
1.4
1.2
1.0
*
Pronoun
Name
Binary Target (he/she)
Nonbinary Target (they)

Table 5a

Prosodic Analysis: Inferential Statistics in Experiment 1

Effect	Estimate (Std. Error)	*t* Value	Pr > |*t*|
Intercept	1.86 (0.03)	56.67	<.0001
Nonbinary vs. Binary	0.05 (0.02)	2.11	0.04
One- vs. Two- Characters	-0.03 (0.02)	-1.31	0.2
Pronoun vs. Name	-0.13 (0.03)	-4.5	<.0001
Nonbinary x One Character	0.02 (0.02)	1	0.33
Nonbinary x Pronoun	0.05 (0.02)	2.06	0.05
One Character x Pronoun	0.04 (0.02)	2.13	0.03
Nonbinary x One-Char. x Pronoun	-0.04 (0.02)	-1.89	0.07

Table 5b

Prosodic Analysis: Estimates of the Nonbinary Effect for Pronouns and Names in Experiment 1

Effect	Estimate (Std. Error)	*t* Value	Pr > |*t*|
Pronoun: Nonbinary Effect	0.2 (0.07)	2.76	0.01
Name: Nonbinary Effect	0.01 (0.06)	0.13	0.90

2.3.4 Analysis #3: Gender Errors

Coders noted whether the response included any incorrect pronouns referring to Alex. For this analysis we included trials excluded for analyses 1 and 2, and only excluded 2 trials that had poor audio. We also analyzed the entire response and not just the critical reference in the subject position. A total of 311 trials in the nonbinary condition and 311 trials in the binary condition were considered for this analysis.

Of the 26 participants, 16 made one or more errors when referring to Alex, in all cases using the pronouns *he/him/his* instead of *they/them/their*. No participant ever used the incorrect gender pronoun for any of the binary characters. There were a total of 29 trials with errors out of 311 stories about Alex, or 9.3%. Table 6 illustrates the different types of errors people made. In

12 of the 29 errors the participant subsequently corrected the error (e.g., "he decided, they decided that they wanted to try out their painting skills and they painted on a canvas"), and in an additional 4 items the participant also used *they* for Alex elsewhere in the response (e.g., "he saw a twenty dollar bill on the ground and decided to pick it up. and they put it in their pocket").

Table 6

Examples and Categorization of Gender Errors in Experiments 1 and 2

Description	Exp. 1 *n*	Exp. 2 *n*	Example
Alex followed by *he/his* and not corrected	8	23	*Alex stopped for lunch with Will at a nearby cafe this afternoon. During the meal...Alex accidentally spilled **his** glass of water.*
Alex followed by *he/his* and corrected	3	3	*Alex stopped for lunch with Will at a nearby cafe this afternoon. During the meal...Alex accidently spilled **h-they**-their water.*
he as subject and corrected	9	2	*Alex had a blast celebrating New Years. At midnight...**he** decided to blow out- **they** decided to blow out candles for the new year and have good wishes.*
he as subject and not corrected	5	3	*Alex visited a floral shop on a sunny afternoon. Right away...**he** bought a ton of flowers for Ana.*
he as subject followed by *they*	1	0	*Alex waited for the subway to arrive one morning. In the station...**he** saw a twenty-dollar bill on the ground and decided to pick it up, and **they** put it in their pocket.*
they followed by *he/him*	3	2	*Alex stopped for lunch with Will at a nearby cafe this afternoon. During the meal…**they** dropped a glass and Will helped him clean it up, helped h-they clean it up.*
Other	0	1	Error reading context sentence "*Alex spent time with his friends*" instead of "*Alex spent time with friends.*"

About half of the errors (15) occurred on the first mention of Alex; the other 14 were on a second mention. Of the later mentions, 9 of the errors (31%) occurred in items that tended to elicit mention of possessives (e.g., "Alex opened his umbrella"; "Alex opened up his suitcase"),

and 5 occurred when the participant elaborated the event in a second clause, e.g. "they dropped a glass and Will helped him clean it up. helped h-they clean it up".

The error analysis demonstrates that about 60% of the participants are still struggling with the use of the nonbinary pronoun. It is notable that the errors consistently misgendered Alex as male. Our illustration of Alex was intended to be androgynous, but perhaps is visually biased toward a male categorization. The name *Alex* is not gender specific but we suspect it is more frequently used for males than females.

One question is whether speakers are less likely to make errors if they have been corrected on a previous error. In this task experimenters only corrected participants if they made an error during the practice trials. Two of the four practice stories presented Alex alone (*Alex took a trip to the theme park last week. During the trip…* [picture shows Alex on a roller coaster]; *Alex sat down on the couch after a long day at school. Then…* [picture shows Alex reading].) 13 of our 26 subjects made one or more misgendering errors on the practice trials (and were corrected), and 13 did not. 11 out of the 13 who got corrected made one or more errors on the critical trials (average error rate = 20%), compared with only 5 of the 13 who didn't make a mistake on the practice items (average error rate = 5%); this difference is significant with a chi-square test ($\chi^2 = 5.85$; $p = .02$.). This suggests that being corrected on a mistake does not increase later success; instead, we observed that people who make mistakes tend to keep making them.

2.4 Discussion

This experiment provided the first experimental evidence about how people produce nonbinary pronouns in a story context. We found that the discourse context (one vs. two people) guided pronoun use similarly for both binary and nonbinary pronouns, leading to greater pronoun use in the one-character condition. This suggests that the conditions for selecting a pronoun vs.

name are applied similarly, consistent with similar evidence for written language (Arnold et al., 2022). This supports the idea that nonbinary *they* is treated as a part of the same pronoun system as binary pronouns.

We also hypothesized that the low frequency of nonbinary *they* might make it harder to produce than binary pronouns *he/she,* and tested how this might affect gender errors and the prosodic prominence of the pronunciation. Results suggested that indeed, nonbinary *they* poses some degree of difficulty for most speakers. About two-thirds of the subjects produced at least one misgendering error in the nonbinary condition, while there were zero errors in the binary condition. On the other hand, importantly, the rate of errors was fairly low. Our analysis of the rate of pronoun production on critical trials excluded any misgendering errors, and we still observed an average of 78% *they* use in the one-person nonbinary condition. All but one of the participants successfully produced *they* on at least one of the critical trials. This suggests that participants were trying to correctly use it.

In addition, there was a tendency to use more perceptually prominent pronunciations for nonbinary *they* than for binary *he* or *she*. While acoustic prominence can sometimes signal differences in information status (Fowler & Housum, 1987; Halliday, 1967), here the information status was identical across binary and nonbinary conditions, suggesting that this was not the reason for prosodic differences. Instead, we draw on evidence that prosodic prominence is also correlated with speech difficulty and disfluency (Arnold & Watson, 2015), which suggests that pronouns were produced less fluently in the nonbinary condition. The prominent pronunciation of *they* may also signal that the use of this form was not "business as usual," but rather was an intentional choice on the part of the speaker.

Our data on the rate of pronoun usage contrasted with findings for written language (Arnold et al., 2022), and demonstrated that our speakers produced nonbinary pronouns just as often as binary pronouns. In fact, they were somewhat more likely to use nonbinary *they* than binary *he* or *she* in the two-character context. This suggests that either participants didn't have difficulty producing *they*, or that difficulty does not always lead to increased name use.

We consider two explanations for the surprising finding that people used pronouns more frequently for the nonbinary than binary referents. One possibility is that for some of the responses in the two-character condition, the participant may have produced "they" and intended the plural interpretation. Even though the target panel clearly showed a single person doing something interesting, participants may have characterized the event more broadly. For example, one picture illustrates Alex or Ana at the supermarket with Liz, where each character is holding their own shopping basket. The critical picture shows Alex or Ana picking up a bottle of milk while Liz stands passively on the other side of the picture. Even though this was meant to illustrate one person buying milk, some subjects may have conceptualized this as a group activity and described it as "they bought milk." If so, the plural use of *they* should have occurred equally in both the binary and nonbinary conditions.

To test this idea, we identified those items that might conceivably elicit a plural interpretation by looking at the binary conditions of the stories for both experiments 1 and 2. A small number of responses in the binary condition used the pronoun *they;* we assumed these pronouns referred to both people (plus one unambiguous "the two of them"), all of which were excluded from analysis for not referring to the target character. We compared the rate of "they" use in the nonbinary condition for these "potentially plural" stories (n=157) compared to the stories where plurals were never produced in the binary condition (n=201), termed "plural

unlikely" stories.[5] For the "plural unlikely" stories, there were numerically more pronouns for the nonbinary (18%) than binary pronouns (13%), but this difference was not significant.

A second possibility is that the experimental context draws attention to the use of nonbinary pronouns. There were several signals to the participants that this experiment was about nonbinary pronoun use. The characters in the stories were explicitly introduced, along with their pronouns, including one character who uses *they/them* pronouns. Our demographic survey also included a question about how many nonbinary people the participant knew. Given the relative infrequency of nonbinary pronouns, both of these would draw participants' attention. This context may have increased usage of nonbinary pronouns specifically, either subconsciously or because of a desire to demonstrate acceptance of nonbinary pronouns.

Experiment 2 sought to replicate the findings from Experiment 1's two-character context, and also tested whether additional exposure to nonbinary *they* increases the likelihood of using it.

3. Experiment 2

Experiment 2 used the same paradigm to further examine pronoun production in a two-person context. Given that this condition elicited few pronouns in Experiment 1, we modified the stories to increase the contextual prominence of the target character. As in Experiment 1, the target was always the subject of the sentence that immediately preceded the response. We increased the prominence of the subject by adding an additional context sentence that mentioned both characters, and by using predicates that provided additional focus on the subject; in most cases these sentences described transfer events where the subject was the goal, since goals are

[5] To identify items that were prone to a plural interpretation, we examined responses for both experiments 1 and 2 identified any item that had at least one plural response in the binary condition, or at least one unambiguously plural response in the nonbinary condition (e.g., "they both"); all of these were excluded from the primary analysis. We then tagged those items as "potentially plural" (n=11), and compared with items that never elicited a plural interpretation (n=13 for exp. 1; n=9 for exp. 2).

particularly likely to be pronominalized (Rosa & Arnold, 2017). We also added a new manipulation to test whether exposing participants to use of the nonbinary pronoun would increase the rate of producing nonbinary *they*.

3.1 Methods

3.1.1 Participants

Twenty-five participants from the University of North Carolina at Chapel Hill participated in exchange for course credit. The data from one nonnative speaker were excluded from analysis. Twenty-four participants are included in the analysis.

3.1.2 Design and Materials

We created 20 new critical stimuli with two characters with the purpose of creating a more constraining semantic context. All the critical items included two characters, but the target character was always in a semantic role that was expected to enhance the focus on the target. In many of the critical stimuli the target was the goal argument in a transfer event (see supplementary material). We manipulated the gender of the target character as in Experiment 1 (nonbinary vs. binary), comparing stories where Alex was the target with stories where Will, Matt, Ana or Liz was the target.

Our second manipulation tested whether exposure to nonbinary *they* would increase use of this pronoun. Participants either were exposed to the nonbinary pronoun *they* or a repeated name *Alex* in six of the filler items that were designated "exposure" items. This was a between-participants manipulation. The exposure stories mentioned Alex twice in the context sentences, and we manipulated whether the second mention was with a pronoun or a name; see Table 7.

Thus, the design was 2 (Nonbinary x Binary) x 2 (Name vs. Pronoun exposure), crossing the target gender and exposure type manipulations. There were four lists such that each had half

nonbinary and half binary critical trials. Two lists included six name exposure trials and two lists included six pronoun exposure trials. On each list there were 20 critical trials, six exposure trials, and 22 filler trials that presented stories with varying numbers of participants and story structures.

Table 7

Experiment 2 Example Stimuli

Critical Stimuli	
Condition	Example
Nonbinary	*Alex visited a floral shop with Liz on a sunny afternoon. Alex borrowed some money from Liz. For Valentine's Day…*
Binary	*Ana visited a floral shop with Liz on a sunny afternoon. Ana borrowed some money from Liz. For Valentine's Day…*
Exposure Stimuli	
Type	Example
Name	*Alex was camping last weekend. Alex went with Liz on a canoe trip. With a splash…*
Pronoun	*Alex was camping last weekend. They went with Liz on a canoe trip. With a splash…*

3.2 Results

3.2.1 Analysis #1: Pronoun or Name?

3.2.1.1 Critical Items. Our primary question was whether responses on the critical items were different by gender condition (nonbinary vs. binary) and priming condition. For this analysis, out of a total possible 480 critical items, 55 items were excluded (11%); see Appendix A.[6]

Figure 5 illustrates that people used pronouns more often in the nonbinary condition, mimicking the findings for the two-person context in Experiment 1. Our model (Table 8) showed

[6] Three coders (the same two as for experiment 1 plus a third) transcribed and coded the responses to the practice, exposure, and critical items. All three coded one subject. 93% of the trials were transcribed the same across all three coders except for minor word choices. Coding decisions were 98% the same across all three.

that the difference between nonbinary and binary conditions was significant, but there was no effect of priming condition nor any interaction between gender and priming.

Figure 5

Experiment 2 Referential Form Analysis Results

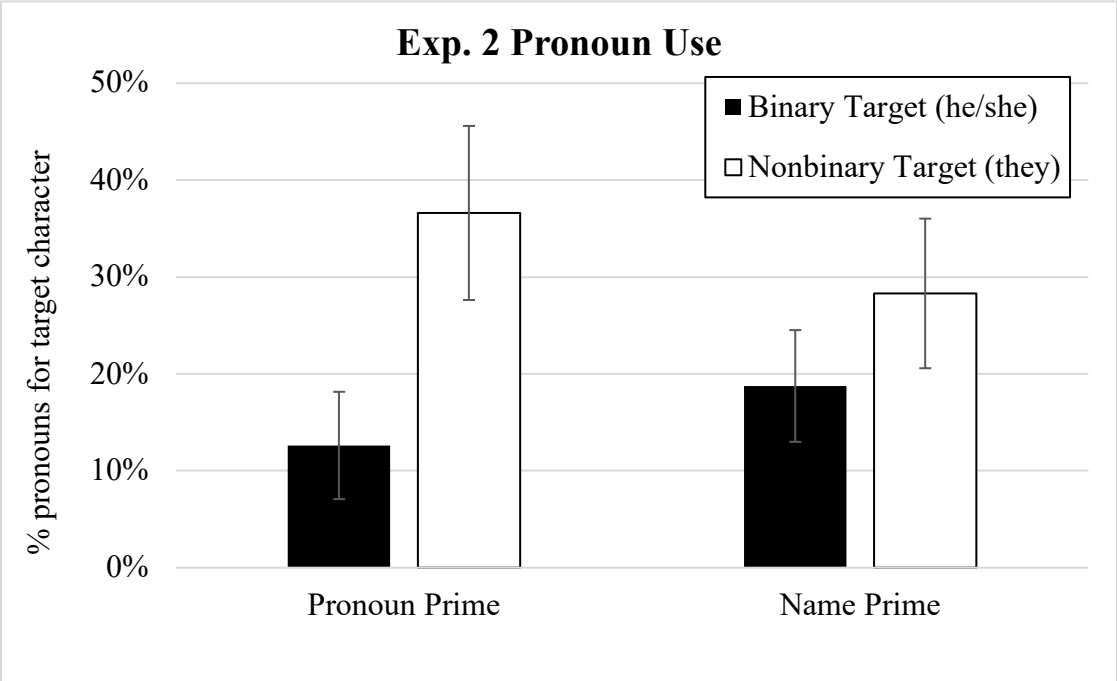


Table 8

Reference Form Analysis: Inferential Statistics for Critical Items for Experiment 2

Effect	Estimate (Std. Error)	*t* Value	Pr > |*t*|
Intercept	-1.49 (0.32)	-4.71	0.0001
Nonbinary vs. Binary Pronoun	0.61 (0.16)	3.7	0.003
Primed vs. Not primed Condition	-0.02 (0.3)	-0.05	0.96
Nonbinary * Primed Character	0.24 (0.15)	1.56	0.13

3.2.1.2 Exposure Items. Our second question was whether the form of the exposure items influenced the participant's choices about how to refer to Alex in their response to that item. We examined responses to the six exposure items. Out of a total of 144 exposure items (six each for

24 participants), 10 trials (7%) were excluded because the participant did not refer to the target in subject position and/or described the event incorrectly.

In the exposure items, participants used pronouns 24% (*SE* = .10) of the time in the name-exposure condition (range 0-83%) and 38% (*SE* = .11) of the time in the pronoun-exposure condition (range 0-100%). We tested the effect of exposure condition using the same analytical approach as in Experiment 1, where this model had one predictor (priming condition), random intercepts for subject and item, and a random slope for priming by item. The effect of priming condition was not significant ($b = 0.5$; $SE = 0.48$; $t = 1.05$; $p = 0.31$).

3.2.2 Analysis #2: Prosodic Prominence as a Signal of Fluency

Again our primary analysis assessed the prosodic prominence of names and pronouns in critical trials, including all the trials in Analysis 1 except for nine trials where the audio was not good enough to hear or where the participant commented before responding or repeated/repaired the target phrase.

As shown in Table 9a and Figure 6, again we found that names were perceived as more prominent than pronouns, with no difference between binary and nonbinary names. Critically, we again found that nonbinary *they* was perceived as more prominent than binary *he* and *she.* Our model supported this pattern: we found a significant difference between pronouns and names, a significant effect of binary vs. nonbinary condition, and an interaction between the two. When we probed the interaction, we found that the nonbinary condition was more prominent than the binary condition for pronouns, but not for names (Table 9b).

We also conducted a secondary analysis, comparing singular and plural productions of "they". As in Exp. 1, the numerical patterns suggested that singular "they" was perceived as more prominent (Avg. = 2.03) than plural "they" (1.8), a difference (0.22) that was comparable

to the difference between binary and nonbinary pronouns (0.25). However, again this effect failed to reach significance (b = 0.10 (SE = 0.05), t = 1.93, p = 0.08).

Figure 6

Average Prominence Ratings for Pronouns and Names in Critical Trials in Experiment 2

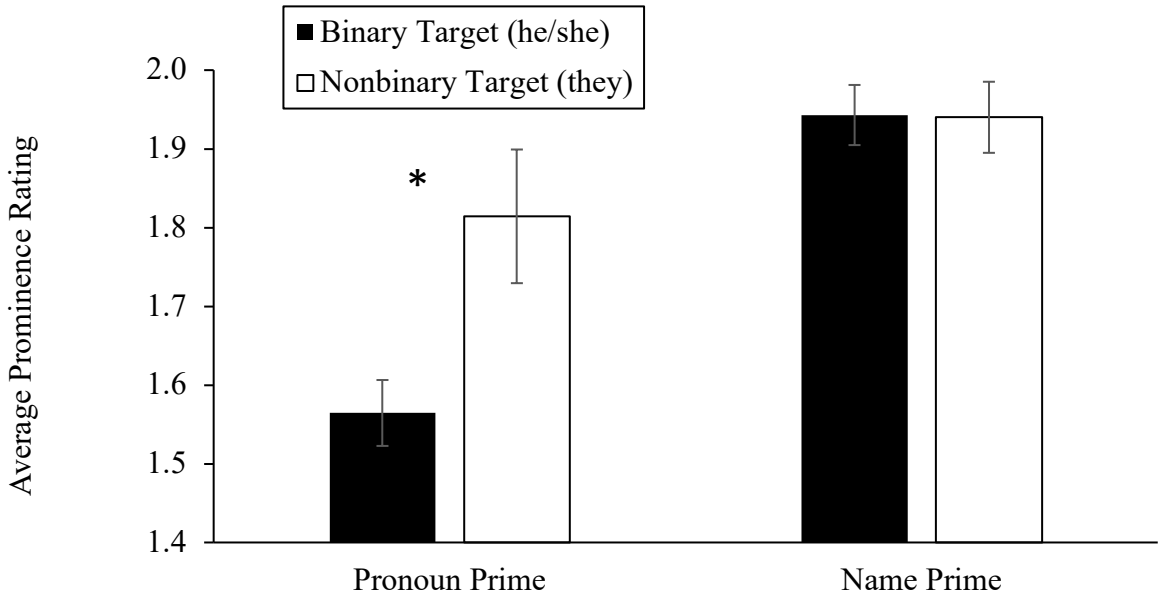


Table 9a

Prosodic Analysis: Inferential Statistics in Experiment 2

Effect	Estimate (Std. Error)	*t* Value	Pr > |*t*|
Intercept	1.82 (0.04)	45.69	<.0001
Nonbinary vs. Binary pronoun	0.06 (0.03)	2.21	0.04
Pronoun prime	0.04 (0.04)	1.07	0.3
Pronoun used	-0.11 (0.03)	-3.16	0.005
Nonbinary x Pronoun prime	-0.02 (0.02)	-1.03	0.31
Nonbinary x Pronoun used	0.06 (0.02)	2.7	0.01
Pronoun prime x Pronoun used	-0.01 (0.03)	-0.2	0.84
Nonbinary x Pronoun prime x Pronoun used	-0.03 (0.02)	-1.13	0.27

Table 9b
Prosodic Analysis: Estimates of the Nonbinary Effect for Pronouns and Names in Experiment 2

Effect	Estimate (St. Error)	*t* Value	Pr > |*t*|
Pronoun: Nonbinary Effect	0.23 (0.08)	2.88	0.01
Name: Nonbinary Effect	-0.01 (0.05)	-0.19	0.85

3.2.3 Analysis #3: Gender Errors

We analyzed the rate of misgendering errors in the nonbinary condition, including all responses to critical and exposure items and not just those that met our inclusion criteria. 19 of 24 participants made one or more error on Alex's pronouns on the exposure or critical items, for a total of 34 errors out of 384 stories about Alex (critical and exposure combined), or 8.9%. Of these, 10 were corrected and 24 were not. There were no errors in the binary condition.

Unlike in Experiment 1, the rate of errors was not related to whether the participant was corrected on the practice trials or not. Out of the 10 people who were corrected for a misgendering error on the practice trials, nine of them made one or more errors on the critical/exposure trials and one did not. Out of the 14 people who were not corrected on the practice trials, 10 made one or more errors on the critical/exposure trials and four did not. This difference was not significant ($\chi^2 = 1.22$, $p = .27$).

The gender errors further underscore the fact that priming had no effect. Participants in the name priming condition made 9% errors on the critical trials and 8% errors on the exposure trials, while participants in the pronoun priming condition made 8% errors on the critical trials and 10% errors on the exposure trials. This means that people made errors even when they had just read *they* out loud from the context sentence, e.g.: *Alex got a new job. They left for work with Will on a rainy day last week. On the way... Alex decided that he-that they would pull out their umbrella for the rainy walk*.

3.4 Discussion

Experiment 2 elicited a numerically greater use of pronouns (103 out of 425, or 25%) than in the two-character condition for Experiment 1 (42 out of 282, or 15%), suggesting that our new stimuli did increase the appropriateness of pronouns for the target character. In this context we replicated several findings from Experiment 1.

First, we again observed that people used nonbinary *they* more than binary *he/she.* Experiment 2 only used two-character contexts, so this confirms the tendency to use *they* more in this context. Again we do not know whether some of these may have been intended as plural, but we estimate that this cannot be the only reason for the difference. As for experiment 1, we examined pronoun use for items deemed "plural possible" vs. "plural unlikely". In experiment 2, the "plural unlikely" items (n=201) elicited pronouns at a greater rate for nonbinary (34%) than binary (16%) pronouns, and this difference was significant ($b = 0.57$ ($SE = 0.24$), $t = 2.38$, $p = 0.036$). This analysis is post-hoc, but it suggests that a plural interpretation cannot account for the entire effect of greater pronoun use in the binary condition.

Experiment 2 also replicated the tendency for nonbinary pronouns to be produced with a more prosodically prominent pronunciation than binary pronouns. In addition, people made gender errors in the nonbinary condition, consistently misgendering Alex as male.

On the other hand, we found no effect of our priming manipulation. We hypothesized that reading "they" used to refer to Alex might increase the rate of using "they" compared to conditions where people read "Alex" to refer to Alex. Notably the priming manipulation was fairly weak, in that it only occurred on six of the fillers. Meanwhile, all of the critical items used repeated names on the second context sentence, either binary or nonbinary. Nevertheless, it is striking that even on our analysis of the exposure items themselves, there was no effect of the

priming condition. Participants produced names over half the time, even in contexts where they had just read "they" referring to Alex (e.g., *Alex needed to practice for an upcoming performance. They recited a poem to Ana on stage. After the performance… Alex took a bow*.)

As in Experiment 1, we hypothesize that the context of the experiment itself may have served as a sort of "global prime" for the use of nonbinary *they*. We know that when nonbinary pronouns are explicitly introduced ("Alex uses they/them pronouns"), it increases the likelihood that comprehenders will interpret *they* as singular (Arnold et. al., 2021). Our experiments created a socially supportive context for using nonbinary *they* by introducing the characters' pronouns. This alone may have drawn attention to this usage.

Indeed, we speculate that this property of our experimental setup explains the greater use of nonbinary than binary pronouns. For most people, nonbinary *they* is so low frequency that it may be generally fairly hard to use, and this may suppress the use of nonbinary pronouns in favor of names (Arnold et al., 2022). But in a social context where nonbinary pronouns are emphasized, people are more likely to use them. This was likely a stronger effect than our priming manipulation.

4. Individual Differences

It seems likely that the ongoing change in how pronouns are used in English is driven by individuals for whom the use of nonbinary pronouns has special personal significance, in particular those whose personal pronouns are *they/them* pronouns or people who are close to those who use them. In support of this, there is evidence that the acceptability of singular *they* for reference to gendered and known individuals is greater for people who are younger and those with greater familiarity with nonbinary gender (Ackerman, 2018; Bradley et. al., 2020; Camilliere et al., 2021; Conrod, 2019).

The current study was not designed to investigate individual differences. The sample size was large enough to examine pronoun usage within the manipulated discourse contexts, but too low to examine additional individual variability. In addition, our participants were all young adults enrolled in Psychology 1, so there was very little age variability. Moreover, we did not probe individual differences in our demographics questions beyond asking for participants' sex assigned at birth, gender identity, and the number of people who they know who identify as nonbinary. None of our participants reported a gender identity other than male or female, and all participants reported the same gender identity as their sex assigned at birth. We also asked participants how many people they know who identify as nonbinary or gender fluid, but there were few people who reported knowing more than 2 (see Table 10).

Table 10. Data on the relation between knowing nonbinary individuals, error rates and the "nonbinary pronoun prominence metric" (average prominence of nonbinary pronouns – average prominence of binary pronouns, by participant, including only those who had data in both conditions, n=40).

How many people do you know who have a nonbinary/gender-fluid identity?	Avg. # errors	Nonbinary pronoun prominence	N
0	1.2	0.15	18
1	1.3	0.53	8
2	1.4	0.27	12
3	1.0	-0.25	2
4	1.3	0.16	4
5	1.2	0.13	6

In an exploratory analysis, combining the data from both experiments, we examined correlations between knowing nonbinary people and our two measures of difficulty producing "they": a) number of errors, and b) nonbinary pronoun prominence. However, the number of

nonbinary people they know was uncorrelated with either errors ($r = -0.01$; $p = 0.96$) or prominence ($r = -0.07$, $p = 0.67$).

Thus, even though exposure to nonbinary *they* is likely to increase fluency with using it, our findings suggest that our participant pool may be too homogenous to detect this effect. All our participants have some degree of familiarity with nonbinary *they*, so everyone was at least moderately successful at using it. However, even those with relatively more familiarity to *they*-users still made mistakes and produced nonbinary *they* with more perceptual prominence than binary *he* and *she.*

5. General Discussion

In two experiments we tested how young adults (college students) use pronouns when telling stories about binary and nonbinary characters. In both experiments we found that for this population, nonbinary pronouns are favored in the same one-character discourse contexts as binary pronouns and tend to be produced at roughly the same rate (or even a little more). This suggests that nonbinary *they* has been subsumed into the same pronoun production framework as binary singular pronouns.

At the same time, we found that participants exhibited speech patterns that signal mild difficulty with using nonbinary *they*. In both experiments there were about 8 or 9% misgendering errors, which in all cases emerged as the use of *he/him/his* for Alex. We never saw any use of an incorrect gender pronoun in the binary condition. In addition, productions of *they* tended to be more prosodically prominent, signaling a less fluent delivery.

In summary, we put college-aged participants in an experiment that required them to talk about a person with personal pronouns *they/them,* and they had some difficulty but were overall fairly successful. These results, together with other evidence from the literature, suggest that any

model of reference production would need to account for the following facts about reference production when referring to people whose personal pronouns are they/them: 1) the discourse context has the same effects on both binary and nonbinary pronoun production (Exps. 1 and 2; Arnold et al. 2022); 2) people make gender errors for nonbinary referents more often than for binary referents (Exps. 1 and 2); 3) in some cases people may favor names for nonbinary referents (Arnold et al., 2022); 4) nonbinary pronouns sound more prominent than binary pronouns (Exps. 1 and 2).

Here we propose a preliminary working model for explaining these findings. We term our model a Usage-Based Model (UBM) of pronoun change because it focuses on language change as something observable in the output. That is, we aim to explain what people actually say, and the meanings derived from those utterances. We assume that a cognitive representation of grammaticality underlies these usages, but such change is not observable unless it leads to changes in pronoun usage. These usages are especially important because they in turn become the input for other people, and the comprehension of these uses may lead to change itself. Note that this model is relevant for speakers that have at least a rudimentary representation of the use of *they*/*them* as a personal pronoun and the existence of gender identities outside the binary. Based on our own experience, we estimate that 10 years ago, most speakers of English did not have this option in their grammar. By contrast, nowadays there is variation across speakers in facility with nonbinary they use (Ackerman, 2018; Bjorkman, 2017; Conrod, 2010; Konnelly & Cowper, 2020).

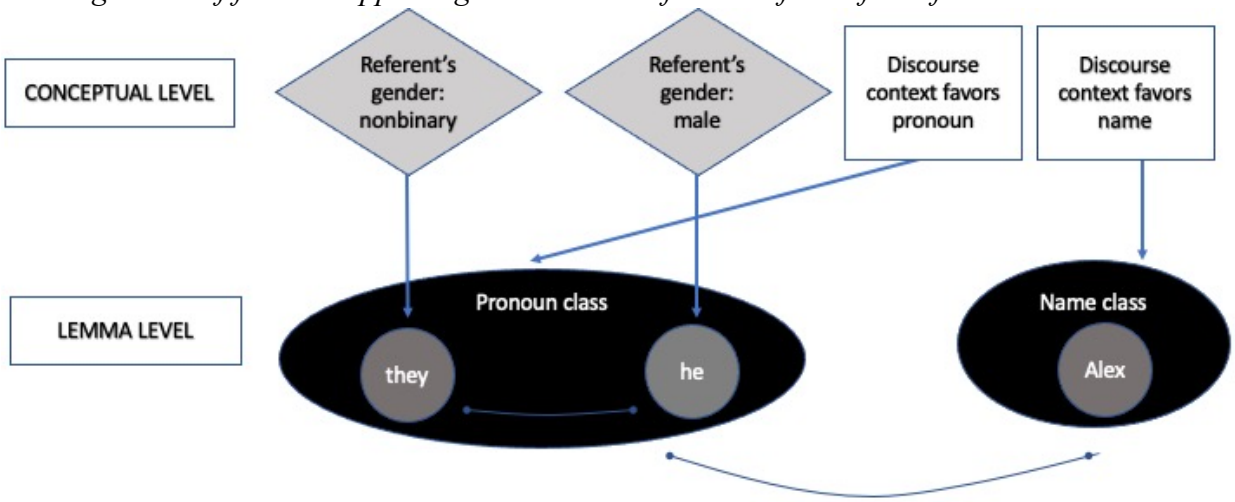
Our working model is based on extensive evidence that speakers use the discourse context to decide on appropriate referential forms, for example using pronouns for referents that are prominent in the discourse context based on how they were treated in the discourse (e.g.,

Ariel, 1990; Chafe, 1976; Gundel et al., 1994). While these models do not discuss psycholinguistic processing, they are consistent with a selection-based model where the discourse context drives referential form choices (Arnold, 2016; Arnold & Zerkle, 2019). One such model has been proposed by Schmitt, Meyer, & Levelt (1999) for reference production in German. German differs from English in that pronoun gender is lexically specified, but we adapt it here for English where gender is instead conceptual.

Schmitt et al. (1999) propose that the discourse context is represented in terms of a binary feature whereby a referent is either 'in focus' or not, and that this feature determines whether a speaker produces a noun phrase (the flower) or a pronoun (it). We follow Schmitt et al. in this simplification, even though other evidence suggests that the discourse context constrains production in a noncategorical fashion, such that some contexts support a relative rather than absolute preference to use pronouns (e.g. Arnold & Griffin, 2007, Kehler et al., 2008, Stevenson et al., 1994; or the data presented here), and the discourse status has a non-categorical and possibly multidimensional nature (e.g., Kaiser & Trueswell, 2008).

Figure 7

Working model of factors supporting alternatives referential forms for reference to Alex



Note. Arrows represent supporting pressures; circle connections represent competition between alternatives.

Figure 7 illustrates a working model, focusing on the constraints most relevant to the production of reference to our character Alex.[7] We hypothesize that the conceptual level contains representations of two critical features. First, the discourse context determines relative appropriateness of different forms. While the details of how it does so are beyond the scope of this paper, our findings show that having one or two people in the story is one constraint. Second, it includes a representation of the referent's gender. For language like English, form choices are driven by the conceptual gender and not lexical gender, although a handful of words may be lexically marked for gender (Ackerman, 2019). Here we present gender in a simplified fashion as "male" and "nonbinary", despite the fact that gender representations are more

[7] Other constraints (such as number or case marking) are not addressed here.

complex (Akerman, 2019). This model also presents a simplified view of the relationship between gender and pronouns; using *they/them* is common for individuals who identify as nonbinary or gender diverse, but it isn't universal. Conversely, some people identify with binary genders but use *they/them* as at least one of their pronouns. Future work is needed to understand the degree to which *they/them* use leads to inferences about gender identity and whether such inferences are accurate; for now the working model assumes at least a probabilistic relation between gender concepts and pronoun use.

We also hypothesize that speakers select a class of reference form as an independent level of representation from the specific word. Here we illustrate this as a class of words at the lemma level. That is, the discourse context determines whether a pronoun is appropriate or not independently from the selection of "she", "he", or "they". Recent findings from a priming paradigm support the hypothesis that speakers activate a broad representation of pronouns as a class (Arnold, 2023). This choice contrasts with other potential expression types such as names, so the choice between pronouns and names are mutually inhibitory. After selecting the pronoun class (similar to Schmitt et al. 1999's gating function), the speaker must select a specific form. These forms are also mutually exclusive, so at the lemma level the specific pronouns are in competition with each other.

This model provides a framework for explaining the four findings listed above. First, this model includes the same discourse context constraints for both binary and nonbinary pronouns. This is consistent with evidence that speakers follow the same discourse constraints for both types of pronouns, as shown by the one vs. two person effect here, and by similar given/new effects in written production (Arnold et al., 2022). Likewise, in comprehension we observe the same bias to assign pronouns to the first-mentioned or subject referent for both binary pronouns

(e.g., Gernsbacher & Hargreaves, 1989; Stevenson et al., 1994) and for nonbinary pronouns (Arnold et al., 2021).

Second, this model suggests that the selection of the pronoun class is driven by the discourse context, but the selection of a particular form is driven by gender at the conceptual level (at least for English). This means it is possible to select the class of pronouns but then make a mistake in selecting the correct form. In principle this could lead to producing *she* for *he* or vice versa, and indeed adults do occasionally make mistakes, but they are rare and were not observed in either experiment. By contrast, there was a consistent but low rate of misgendering errors for reference to Alex in both experiments. We hypothesize that this stems from the relative strength of representations at both conceptual and lemma levels.

At the conceptual level, the representations of binary male/female genders are likely stronger than the representation of nonbinary gender in the abstract sense due to people's greater experience with binary genders (e.g., Akerman, 2019). This abstract gender representation likely modulates the strength of the representation for Alex as an individual, such that the nonbinary representation may be weak. In addition, the dominance of binary gender in our language and our social world may lead to the automatic partial activation of a binary gender for all characters, including Alex. Given that we only saw misgendering errors with male pronouns, we assume that some of our participants considered Alex to have male characteristics. In real world interactions, people may also have trouble remembering people's personal pronouns when they don't match their expectations. Gardner and Brown-Schmidt (2023) presented participants with vignettes about fictional characters, and found that participants only remembered that a character used *they/them* pronouns about 50% of the time.

At the lemma level, the use of *they* as a singular pronoun for specific referents is relatively low frequency, so production may be slowed compared to the production of *he* and *she* (Griffin & Bock, 1998; Jescheniak and Schriefers 1994). In addition, the activation of *they* for a specific referent suffers from competition with the more frequent binary forms. Research on word choice suggests that word alternatives compete with each other (e.g. Dell, 1986; Britt et. al., 2016; Griffin & Bock, 1998; Jescheniak & Levelt, 1994), and words with similar meaning are often both activated during production (Jescheniak & Schriefers, 1998; Peterson & Savoy, 1998). In this context, the weak activation of *they* makes it more susceptible to competition from gendered pronouns *he* or *she*. Given the potential for activation from a male representation of Alex, sometimes *he* is activated more than *they.*

On the other hand, there were also several aspects of our task that may have promoted the use of *they*. We explicitly introduced Alex's pronouns, making them salient in the context. The instructions also included a test to make sure participants knew each character's name and pronouns. Research suggests that even just one usage of *they* in reference to a person is enough to dramatically increase use of *they*. Kramer et al. (2022) examined pronoun use in an experimental task where participants wrote narratives about pictured individuals who presented as feminine, masculine, or androgynous. Participants were much more likely to use *they* when referring to the androgynous-presenting characters when they had previously read the pronoun *they* referring to them (77%) than if they had not (8.6%). This suggests that our experimental task alone may have supported the use of nonbinary *they.*

In the two experiments reported here, we did not find any suppression of pronouns for Alex in favor of using the name. However, in another study Arnold et al. (2022) found that writers were less likely to use pronouns for referents whose personal pronouns are *they/them* than

for referents whose personal pronouns are *he/him* or *she/her*. This finding could also be explained by this model, in that pronouns and names compete. If the target pronoun *they* is not fully activated, or if it is competing with *he*, the activation for using a pronoun will be lower and the name may be more likely to be selected.

The competitive nature of this model also accounts for our finding that the pronunciation of nonbinary *they* was more prominent than the pronunciation of binary pronouns *he* and *she*. If *they* competes with *he* for both conceptual and lemma-frequency reasons, this delays the selection of the word. When the context supports a lower-frequency or less accessible form, production can be delayed and word pronunciations are more prominent (Arnold, Kahn, & Pancani, 2011; Bell et al., 2003; Ferreira & Swets, 2002; Kahn & Arnold, 2010; 2015; Watson, Arnold, & Tanenhaus 2008).

This model also makes it clear that form choices are driven by multiple simultaneous factors, including features of the current situation. Our experimental task was designed to make nonbinary *they* contextually acceptable (e.g. introducing Alex's pronouns) because this is the only appropriate context in which to test its usage. This may explain two of our findings. First, in experiment 2 we found that priming nonbinary *they* in the filler contexts had no effect. It may be that the instructions themselves were strong enough to focus attention on nonbinary *they*, and on top of this there was no additional effect of priming.

Second, speakers unexpectedly used *they* somewhat more than *he/she* in the two-person context. We considered whether participants might have interpreted the depicted actions as having plural actors instead of a singular target actor. While this may have occurred for some items, we found that this pattern also occurred even for stories that were never described as a plural action in the binary condition. We therefore instead speculate that *they* use increased

because the use of nonbinary pronouns was salient for our experimental task. Our student participants may have wished to demonstrate successful use of Alex's appropriate pronouns, leading to an increase in nonbinary *they*. The one-person context already strongly supported the use of both binary and nonbinary pronouns, leaving less room for this "contextual activation" effect.

Our model also provides a framework for speculating about how nonbinary *they* use is changing over time. Change can be represented as the frequency with which people produce nonbinary *they* in an appropriate context, or as the fluency and accuracy with which they do so. The observance of any tokens at all depends on contexts in which there is an individual with the personal pronouns of *they/them.* We hypothesize that this context is critical because this particular usage is different from other singular uses of *they/them* (e.g., those occurring in Konnelly & Cowper's Stage 1 and 2), many of which have been around for centuries. The use of *they/them* as one's personal pronouns frequently (although not always) signals a gender-nonbinary or gender-fluid identity (Sanders, 2019). Thus, language change is inextricably tied to changing concepts of gender.

In a specific instance of referring, as in our storytelling task, the selection of a pronoun or name is driven by several contextual pressures. As with all referring situations, the discourse context strongly influences the appropriateness of pronouns or names. In addition, the use of nonbinary *they* is influenced by two additional pressures: a) social pressure to use or not use this form, and b) variation in familiarity with the linguistic and conceptual representations supporting its usage.

As Conrod (2018) and Konnelly & Cowper (2020) point out, the recent adoption of nonbinary *they* is a change "from above", meaning that speakers are consciously aware of the

change (Labov, 1966). The use of this form is far from neutral and carries social significance. Some of this significance comes from individuals who advocate for the use of inclusive and respectful language at a personal level. For example, many professors list their pronouns on syllabi, and many faculty and staff list their pronouns in email signatures. Some students publicly announce that their personal pronouns are *they/them* in classes. Institutional policies also impact perceptions about the type of language that is expected, especially in public situations like classrooms, meetings, and written policies. In April 2023, UNC's official policy was to provide "an inclusive and welcoming environment for all members of our community. Consistent with that commitment, gender-inclusive terms (chair; first-year student; upper-level student, etc.) should be used on University documents, websites and policies." (Policy on Gender-Inclusive language, 2023). In other contexts, transphobic attitudes may have the inverse effect, discouraging the use of *they* to refer to specific individuals (Conrod, 2018). The experiment itself presented a discourse situation of public language; the interviewers were students and thus peers of the participants, and the mere fact that they introduced our characters' pronouns signaled that the use of nonbinary *they* was valued in this situation.

Independent of political attitudes about *they/them* pronouns, participants also vary in their exposure to both *they/them* pronouns and to nonbinary gender identities. Both of these pressures are hypothesized to modulate the strength of the representations in our model. Thus, in a specific referring situation, the lemma *they* may be more or less available as a function of how frequently the speaker has used the word in the past. In addition, the use of *they* as a personal pronoun is probabilistically associated with a nonbinary gender identity. People with little exposure to this concept may have a strongly binary representation of gender. If so, they may automatically categorize the referent as either male or female; in the case of our experiment many participants

appeared to view Alex as male. Thus, participants may only variably or incompletely activate a nonbinary gender representation for Alex. If the competing "male" representation is activated, this will instead increase activation on the competing pronoun *he*.

In sum, real-life productions of nonbinary *they* are dependent on numerous constraints, including the appropriate discourse context, political attitudes and/or social pressure to use or avoid nonbinary pronouns, gender concepts, and situation-specific support for particular forms. This process itself is critical because it results in variable output, and this output itself influences future references. In situations where *they* is socially promoted, people feel compelled to try to use it, even if they have not habitually done so in the past. In public contexts like classes or in written documents, this normalizes the use of *they.* It also provides the input into other people's grammatical systems. According to MacDonald's PDC framework (MacDonald, 2013), cognitive constraints on production drive the frequency of linguistic forms, and this frequency in turn drives the development of linguistic knowledge and comprehension processes. Thus, each instance of referring has the potential to impact the cognitive status of nonbinary *they* for both the speaker and their addressees.

This model also provides a framework for thinking about the role of speaker intention in the integration of nonbinary *they* into mainstream discourse. Much work on psycholinguistics focuses on the automatic processes that occur as words or concepts are activated (e.g., Swinney, 1979). But it is also well known that language production involves monitoring (e.g., Levelt, 1989), and speakers can inhibit the production of activated phrases that are taboo (Motley, Camden & Baars, 1982). Patterns of *they* production are undoubtedly driven by the speaker's intentional selection of *they* in the face of the automatic activation of other pronouns, for example *he* in our experiments. Thus, the speaker's intention to produce respectful pronouns

plays a critical role in this ongoing change. This is not a value-neutral choice, given that the health of transgender and gender-diverse individuals is tied to respectful pronoun usage (Sevelius et al., 2020). Thus, every speaker has the opportunity to make a difference, one pronoun at a time.

6. Data Accessibility

Data and supporting materials for this paper are available at https://osf.io/c2j65/ and https://arnoldlab.web.unc.edu/publications/supporting-materials/supporting-material-for-arnold-venkatesh-vig/.

7. Ethics and consent

This study was approved by the Institutional Review Board at the University of North Carolina at Chapel Hill (protocol # 21-2784), and all participants provided informed consent.

8. Acknowledgements

This research was partially supported by NSF grants 1917840 to J. Arnold. Thank you to A'sjei Scott for help with running subjects, coding, and analysis. Thank you to Gabrielle Garner, Nicholas Payst, Eri Kakoki and Avery Wall for their help with the prosodic analyses. The faces for Liz, Will, and Alex were drawn by Darith Klibanow and are copyrighted to Jennifer Arnold 2020. The faces for Ana and Matt were drawn by Eri Kakoki and are copyrighted to Jennifer Arnold 2022.

9. Competing Interests

The authors have no competing interests to declare.

10. Authors' contributions

J. Arnold: Concetualization, Methodology, Validation, Formal analysis, Resources, Data Curation, Writing – Original Draft; Writing – Review & Editing, Visualization, Supervision, Project administration, Funding acquisition

R. Venkatesh and Z. Vig: Methodology, Software, Investigation, Data Curation, Writing – Review & Editing, Visualization

11. References

Ackerman, L. (2018). Our words matter: acceptability, grammaticality, and ethics of research on singular 'they'-type pronouns. PsyArXiv 10.31234/osf.io/7nqya

Ackerman, L. (2019). Syntactic and cognitive issues in investigating gendered coreference. *Glossa*: *a journal of general linguistics, 4*(1), Article 117. https://doi.org/10.5334/gjgl.721

Ariel, M. (1990). *Accessing noun-phrase antecedents*. London: Routledge.

Arnold, J. E. (2023). Hearing pronouns primes speakers to use pronouns. Talk, Human Sentence Processing Conference, Pittsburgh, PA

Arnold J. & Griffin Z.M. (2007). The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language, 56*(4), 521-536. https://doi.org/10.1016%2Fj.jml.2006.09.007

Arnold, J. E. (2016). Explicit and emergent mechanisms of information status. *TopICS*, 8, 722-736 DOI: 10.1111/tops.12220

Arnold, J. E. (1998). *Reference Form and Discourse Patterns* [Unpublished doctoral dissertation]. Stanford University.

Arnold, J. E. & Watson, D. G. (2015). Synthesizing meaning and processing approaches to prosody: performance matters. *Language, Cognition, and Neuroscience, 30*(1-2), 88-102. http://dx.doi.org/10.1080/01690965.2013.840733

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

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

Arnold, J.E., Kahn, J.M., & Pancani, G. (2012). Audience design affects acoustic reduction via production facilitation. *Psychonomic Bulletin and Review*, 19, 505-512. DOI 10.3758/s13423-012-0233-y

Arnold, J. E., Rosa, E. C., Klinger, M., Powell, P., & Meyer, A. (March 2014) Mechanisms of prosody production: Differences between children with and without ASD. Poster, CUNY conference on human sentence processing, Columbus, OH.

Baron, D. (2020). *What's your pronoun? Beyond He and She*. Liveright Publishing.

Bate, E. (2021). Demi Lovato Announced They Are Nonbinary In A "Vulnerable" Instagram Post. Buzzfeed news. https://www.buzzfeednews.com/article/eleanorbate/demi-lovato-non-binary-pronouns.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C.Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America, 113*, 1001–1024. doi:10.1121/1.1534836

Ben, F. (July 17, 2019). It's Personal: Is 'They' the Way to Go? Letters to the Editor Re "The Perfect Pronoun: Singular 'They,'" by Farhad Manjoo (column, July 11). NY Times.

https://www.nytimes.com/2019/07/17/opinion/letters/grammar-gender.html. Downloaded June 21, 2021.

Bjorkman, B., (2017). Singular *they* and the syntactic representation of gender in English. *Glossa: a journal of general linguistics*, *2*(1), Article 80. https://doi.org/10.5334/gjgl.374

Bradley, E. D., Salkind, J., Moore, A., & Teitsort, S. (2019). Singular 'they' and novel pronouns: Gender-neutral, nonbinary, or both? *Proceedings of the Linguistic Society of America, 4*, Article 36. https://doi.org/10.3765/plsa.v4i1.4542

Britt, A. E., Ferrara, C., & Mirman, D. (2016). Distinct effects of lexical and semantic competition during picture naming in younger adults, older adults, and people with aphasia. *Frontiers in Psychology*, *7*. DOI: https://doi.org/10.3389/fpsyg.2016.00813

Camilliere, S., Izes, A., Levanthal, O., & Grodner D. (2021). *They* is changing: Pragmatic and grammatical factors that license singular *they*. *Proceedings of the Annual Meeting of the Cognitive Science Society, 43*. https://escholarship.org/uc/item/3tc9s9b0

Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N. Li (Ed.), *Subject and topic*, (pp. 25-56). New York: Academic Press Inc.

Conrod, K. (2019). *Pronouns Raising and Emerging* [Unpublished doctoral dissertation]. University of Washington.

Conrod, K. (2020). Pronouns and Gender in Language. In K. Hall and E. Barrett (Eds.), *The Oxford Handbook of Language and Sexuality*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190212926.013.63

Conrod, Kirby (2018, Oct. 21). Pronouns and Misgendering [Conference presentation]. New Ways of Analyzing Variation (NWAV) 47, New York University, New York, NY, United States.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*, 283-321.

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

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language, 46*, 57–84. doi:10.1006/jmla.2001.2797

Foertsch, J. & Gernsbacher, M. (1997). In search of gender neutrality: 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

Fowler, C., & Housum, J. (1987). Talkers' signalling of 'new' and 'old' words in speech and listeners' perception and use of the distinction. Journal of Memory and Language, 26, 489–504.

Gardner, B. and Brown-Schmidt, S. (2023). Improving Memory for and Production of Singular They Pronouns. Manuscript, U. Vanderbilt.

Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language*, *38*, 313– 338. DOI: https://doi.org/10.1006/jmla.1997.2547

Gross, E. B., Kattari, S. K., Wilcox, R., Ernst, S., Steel, M., & Parrish, D. (2022). Intricate Realities: Mental Health among Trans, Nonbinary, and Gender Diverse College Students. *Youth*, *2*(4), 733–745.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring

expressions in discourse. *Language, 69*, 274-307.

Halliday, M. A. K. (1967). Intonation and grammar in British English. The Hague: Mouton.

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, & Cognition*, *20*, 824–843. DOI: https://doi.org/10.1037/0278-7393.20.4.824

Jescheniak, J. D., & Schriefers, H. (1998). Discrete serial versus cascaded processing in lexical access in speech production: Further evidence from the coactivation of near-synonyms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1256–1274. DOI: https://doi. org/10.1037/0278-7393.24.5.1256

Kahn, J. M., & Arnold, J. E. (2012). A processing-centered look at the contribution of givenness to durational reduction. *Journal of Memory and Language*, *67*(3), 311–325. https://doi-org.libproxy.lib.unc.edu/10.1016/j.jml.2012.07.002

Kaiser, E., & Trueswell, J. C. (2008). Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference resolution. *Language and Cognitive Processes, 23,* 709–748.

Konnelly, L. & Cowper, E. (2020). Gender diversity and morphosyntax: An account of singular *they. Glossa: a journal of general linguistics*, *5*(1), 40. https://doi.org/10.5334/gjgl.1000

Kramer, M. A., Boland, J., & Queen, R. (2022). Getting to know them: The influence of familiarity on the production of singular specific *they* [Unpublished manuscript]. https://doi.org/10.31234/osf.io/v5yjz

Labov, William. 1966. *The social stratification of English in New York City.* Washington, DC: Center for Applied Linguistics.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.

Leventhal, O., Camilliere, S., Chen, Peiyao, & Grodner, D. (2020, Mar. 20). Using ERPs to investigate the processing of singular *they* [Poster presentation]. CUNY Conference on Human Sentence Processing, Amherst, MA, United States (virtual).

MacDonald, M. C. (2013). How language production shapes language form and comprehension. Frontiers in Psychology, 4, 226. https://doi.org/10.3389/fpsyg.2013.00226

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

McWhorter, J. (2018, Sep. 4). Call them what they wants. *The Atlantic*. https://www.theatlantic.com/ideas/archive/2018/09/the-new-they/568993/

Motley, M. T., Camden, C. T., & Baars, B. J. (1982). Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning & Verbal Behavior, 21*(5), 578–594. https://doi.org/10.1016/S0022-5371(82)90791-5

Moulton, K., Block, T., Gendron, H., Storoshenko, D., Weir, J., Williamson, S., & Han, C., (2022). Bound Variable Singular They Is Underspecified: The Case of All vs. Every. *Frontiers, 13*. doi: 10.3389/fpsyg.2022.880687

Nunberg, G. (2016, Jan. 13). *Everyone uses singular 'they,' whether they realize it or not* [Radio broadcast]. Fresh Air. https://www.npr.org/2016/01/13/462906419/everyone-uses-singular-they-whether-they-realize-it-or-not

Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 539–557. DOI: https://doi.org/10.1037/0278-7393.24.3.539

Policy on Gender-Inclusive language. (2023). https://policies.unc.edu/TDClient/2833/Portal/KB/ArticleDet?ID=132161

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

Prasad, G., & Morris, J. (2020). The P600 for singular "they": How the brain reacts when John decides

Rosa, E. C., & Arnold, J. E. (2017). Predictability affects production: Thematic roles can affect reference form selection. *Journal of Memory and Language, 94*, 43-60. https://doi.org/10.1016/j.jml.2016.07.007

Sanders, W. (2019). What People Get Wrong About They/Them Pronouns. *them*. https://www.them.us/story/coming-out-they-them-pronouns

Sevelius J.M., Chakravarty D., Dilworth S.E., Rebchook G., & Neilands T.B. (2020). Gender affirmation through correct pronoun usage: Development and validation of the Transgender Women's Importance of Pronouns (TW-IP) Scale. *International Journal of Environmental Research and Public Health*, *17*(24), Article 9525. https://doi.org/10.3390/ijerph17249525

Sevelius JM, Chakravarty D, Dilworth SE, Rebchook G, Neilands TB. (2020). Gender Affirmation through Correct Pronoun Usage: Development and Validation of the

Transgender Women's Importance of Pronouns (TW-IP) Scale. Int J Environ Res Public Health, 19;17(24):9525.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18, 645-659.

to treat themselves to sushi. https://doi.org/10.31234/osf.io/hwzke

Watson, D. G., Arnold, J. E., & Tanenhaus, M. K. (2008). Tic Tac TOE: Effects of predictability and importance on acoustic prominence in language production. *Cognition, 106*(3), 1548-1557. doi:10.1016/j.cognition.2007.06.009

Zerkle, S. & Arnold, J. E. (2019). Does pre-planning explain why predictability affects reference production. *Discourse and Dialogue, 10*(2), 34-55. https://doi.org/10.5087/dad.2019.202

Appendix A

Excluded Trials for the Reference Form and Prosodic Analyses for Experiments 1 and 2

	Experiment 1		**Experiment 2**		
			Ref. Form Analysis		
	Ref. Form Analysis	**Prosodic Analysis**	**Critical**	**Exposure**	**Prosodic Analysis**
Subject is not target or wrong event	24	24	34	10	34
Wrong name, wrong pronoun, or correction on referring expression	22	22	2	0	2
Context incorrect or didn't read prompt or changed structure	3	3	15	0	15
Audio problems	2	2	0	0	0
No subject NP	2	2	4	0	4
Repeated/repaired target	*N/A*	11	*N/A*	*N/A*	7
Can't hear well enough to code prominence	*N/A*	4	*N/A*	*N/A*	1
Commented before response	*N/A*	0	*N/A*	*N/A*	1
Total in Analysis	571	556	425	134	416
N Subjects	26	26	24	24	24
N Items per Subject	24	25	20	6	20
% Excluded	8	11	11	7	13

Appendix B

We analyzed perceptual prominence using a coding system designed for an earlier experiment (Arnold et al., 2014). This coding system asks listeners to distinguish between 3 broad categories (unstressed, somewhat prominent, and very prominent) with the codings 1, 2, and 3. To allow coders to recognize finer-grained distinctions we also included 1.5, 2.5, and 3.5. While participants may use these numbers differently in an absolute sense, the average of multiple codings should reveal relative differences across conditions.

For each analysis, four undergraduate research assistants (total 6 people) listened to the participant responses and used the following instructions to code the degree to which the critical word (name or pronoun) sounded prominent within the sentence; our final data was the average of the four coders. This approach adapts the technique of using naïve perceptual coding (Cole, Mo, & Hasegawa, 2010; Cole, Mahrt, & Roy, 2017) with four changes. First, our coders were not completely naïve, although they were not trained phoneticians. Second, our 3-point rating scale was more fine-grained than Cole et al.'s categorical distinction between prominent and not prominent. Third, we asked our coders to rate all the tokens instead of just a subset, which provides greater reliability in the comparison across conditions. Fourth, we used fewer coders than Cole et al. (2017), who found that for prominence coding required a minimum of 5 naïve coders for stable measures. However, the loss in granularity by having fewer coders was offset by the increase in consistency by asking our coders to rate all the items.

In the primary analysis, coders listened to the target sentence in the same audio file that contained the context sentence, so it is possible they may have also listened to the context sentence (although they were not instructed to do so). In the secondary analysis we compared the critical trials that used nonbinary "they" with filler items where "they" was used in a plural sense, only including participants who had data in both conditions. For coding these items, the context sentence was removed, so coders could not easily distinguish the singular from plural conditions. In the primary analysis we collected codings from four people for each experiment for the critical items (For Exp. 1: ZV; EK, NP, & AW; for Exp. 2: RV, ZV, NP, & AW). In the secondary analysis the four coders were RV, ZV, NP & GW for both experiments.

The two sets of codings both included the subset of critical trials where speakers used "they" for a nonbinary target. This offers an opportunity to assess the data for consistency across the two analysis sets. For experiment 1, the average ratings for analysis 1 and analysis 2 were correlated at $r = 0.78$, and for experiment 2, they correlated at $r = 0.83$. This relatively high correlation supports the reliability of these perceptual codings.

Instructions for Coding Prosodic Prominence

Go through each file and listen to the response, focusing on the critical name or pronoun in the response. You don't have to listen to the context sentences, only to the final response sentence beginning with the prompt (e.g., "Suddenly…" or "At that time…"). Listen to how prominent/emphatic the pronunciation sounds and code it on the following scale:

3.5 = exceptionally emphatic, more contrastive than usual
3 = really prominent and emphatic, even contrastive sounding: the <u>PANDA</u> spins
2.5 prominent-sounding (accented), but less than 3
2 = somewhat prominent, but not strongly accented: the PANDA spins

1.5 – de-stressed and backgrounded, but not as much as it could be
1 = de-stressed, sounds backgrounded: the panda spins, or the panda SPINS

You can also use .5 markings to indicate levels that fall between these three: 1, 1.5., 2, 2.5, 3, 3.5

Note that the verb (spins) can vary in how it is pronounced too. Sometimes both words might sound prominent. The verb's pronunciation may affect how you hear the target word, which is expected. However, your rating is for how prominent the target is in an absolute sense, and not strictly in relation to the verb (despite the fact that the verb may push around your perception).

Note: these examples are given in terms of a carrier sentence "the Panda spins" that is not used in this experiment, but the same idea applies to all the names and pronouns.