Processing cataphoric *they* amidst pronominal innovation

Steven Foley* (University of Southern California) & Byron Ahn (Princeton)

*foleys@usc.edu

Draft: January 25, 2024

Abstract

When they encounter a cataphoric pronoun during real-time sentence processing, the comprehender begins searching actively for a feature-matched noun that can supply its reference. The present study investigates individual variation in this active search procedure, leveraging an ongoing change in the pronoun system of North American English. The types of referents compatible with the *they*-series of pronouns is expanding, with an increasing number of speakers allowing definite singular referents: especially but not exclusively when referring to someone with a nonbinary gender identity. Sociolinguistic work shows that the speakers who most accept innovative usages of singular *they* tend to be younger, or to be non-cisgender (e.g. transgender, nonbinary). Recruiting participants representing diverse ages and gender identities, the present experiment tracked reading times of sentences involving cataphora. Results shows that cataphoric *they* is processed differently than cataphoric *s/he*, across the board. There is a significant processing cost to reading a plural noun that follows cataphoric *s/he*, indicating that singular cataphors evoke strong number-expectations. However, the cost of reading a singular noun after *they* is smaller and emerges later; *they* seems to evoke weak number-expectations. Individual differences show that those expectations are especially weak for younger participants, perhaps reflecting a higher baseline familiarity with singular *they*. On the other hand, course-grained gender identity (cis vs. non-cis) does not seem to be a reliable predictor of how cataphoric *they* is processed. Integrating insights from sociolinguistics and psycholinguistics, this study offers a novel view into ongoing language change and its manifestation in real-time processing measures.

Keywords: sentence comprehension, cataphora, grammatical number, singular *they*, individual differences, age, gender identity

1 Introduction

Gaining usage in North American English is so-called singular *they* (Bjorkman 2017, Konnelly and Cowper 2020, Conrod 2022): an umbrella term for several usages of that pronoun, including some that are more recent and innovative than its prevalent plural use.[1] Certain "definite specific" usages

---

[1]Throughout, an italicized nominative-case pronoun is a shorthand for any morphological form from that pronoun-series: *they* = *they/them/their/theirs*, *he* = *he/him/his*, *she* = *she/her/hers*, etc.

of singular *they* are particularly characteristic of younger individuals, and those with transgender or nonbinary gender identities — as shown by off-line measures like acceptability ratings (Conrod 2019, Camilliere et al. 2021). Building on this observation, our study addresses a psycholinguistic question and a sociolinguistic one. To what extent do users of contemporary English expect *they* to have a singular referent, rather than a plural one, when actively comprehending real-time linguistic input? And, do those social variables that have been shown to correlate with off-line measures also predict individual variation in on-line measures like reading times?

In addressing the sociolinguistic question, we recruited speakers from a range of ages and gender identities, as participants in a reading-time experiment. The design of that study addressed the psycholinguistic question. Key stimuli involved sentences with cataphora, also known as backwards anaphora. This linguistic relation occurs when a pronoun precedes the noun phrase that supplies its reference, the postcedent.[2] Previous work in sentence processing has shown that cataphora is processed actively. That is, upon encountering a pronoun in a potentially cataphoric position, the comprehender will actively anticipate that a corresponding postcedent noun will appear in the nearest upcoming position that is grammatical. The nature of that linguistic expectation should depend on features of the cataphoric pronoun, and the comprehender's experience with the usages of that pronoun.

Results of our experiment show that there is an especially strong expectation for a singular postcedent after reading cataphoric *he* or *she*, among all participants. But after cataphoric *they*, the parallel expectation for a plural postcedent is, at best, weaker. It is particularly weak among younger participants, suggesting that age is a good predictor of an individual's number-expectations for *they*; the older, the stronger the expectation for plural. Gender identity, on the other hand, seems to be a noisier proxy for singular-*they* expectations: on average, our cisgender participants did not process cataphoric *they* in a way that is reliably different than our non-cisgender participants did — even though the transgender and nonbinary speakers tend to rate singular *they* better, independently of age. So, at least given the present experimental methodology and design, not every social variable that correlates with off-line acceptability of singular *they* also predicts differences in the real-time processing of *they*.

Synthesizing insights from sociolinguistics and psycholinguistics, this study innovates on previous work in a few ways. Most experimental work on singular *they* has involved off-line judgement tasks, investigated anaphoric dependencies, and recruited from relatively homogeneous participant populations. In contrast, we use a reading-time methodology that tracks real-time sentence comprehension; on-line measures help abstract away from individuals' conscious awareness of singular *they*, a linguistic phenomenon that is socially charged in contemporary North America. We also investigate cataphora rather than anaphora, since a pronoun-first configuration requires comprehenders to make predictions about the interpretation of a pronoun, before ever encountering the noun or name supplying its reference. And, rather than recruiting mostly cisgender undergraduates, our participants represent a range of generations and gender identities, offering a more diverse snapshot of sociolinguistic variation.

The rest of the paper is structured as follows. Section 2 provides relevant linguistic background information. Section 3 details the experimental design and reports the results. Discussion of the

---

[2]Throughout this paper, when "(co)refer", "(co)reference", or "(co)referent" is used when describing a pronoun's interpretation, these words are used with their commonsense meaning. More technically, not all pronouns refer or have a referent; some are "bound variables" whose interpretation covaries with its post-/ante-cedent. We return to this in §2.

findings is in Section 4, and Section 5 concludes.

2   Background

To contextualize our research questions and experimental design, this section provides some background on the linguistic properties of English pronouns (Section 2.1) and on the processing of cataphora (Section 2.2). Finally, some linking hypotheses relevant for interpreting the experimental results are presented (Section 2.3).

2.1   Third-person pronouns in English

Pronouns get their interpretation in a number of ways. A pronoun might refer to an individual named by a non-pronominal referential expression (noun or proper name), or it might be interpreted as a variable that is quantificationally bound. The following data illustrate; subscript numbers indicate the intended coreference relations. In the referential case (1), the pronoun points to a specific individual named by a definite noun phrase, '*the actress*'. In the bound-variable example (2), the pronoun covaries in interpretation with each member of the relevant set of actresses.

(1)   [The actress]$_1$ stubbed her$_1$ toe.   **Referential *she***

(2)   [Each actress]$_1$ stubbed her$_1$ toe.   **Bound-variable *she***

Across languages, pronouns are subject to various interpretive constraints. Some (e.g. Principles A/B/C of Binding Theory; Chomsky 1981, et seq.) apply to any pronoun in a particular syntactic position. Other constraints are dependent on grammatical features inherent to the pronouns themselves. In English, for instance, gender–animacy features distinguish *he* (masculine), *she* (feminine), and *it* (inanimate); number features distinguish those three pronouns (singular) from *they* (plural). Note that number features in this sense are diagnosed morphosyntactically, by verbal agreement. When the subject is either *he, she, it*, or a morphologically singular noun phrase, there is one form of the verb (3a); there is another form (3b) when the subject is the pronoun *they* (regardless of how many individuals it refers to) or a plural noun phrase.[3]

(3)   a.   { He / She / It / The teacher } <u>has</u> arrived, <u>was</u> arriving, <u>arrives</u>.

    b.   { They / The teachers } <u>have</u> arrived, <u>were</u> arriving, <u>arrive</u>.

In general, pronouns and noun phrases intended to be coreferential are required to accord in linguistic features as much as possible (e.g., Heim 2008). So, it is generally infelicitous to use the pronoun *it* (inanimate) to refer to a human individual (e.g., *the geographer*), or to use *he* (masculine) to refer to a feminine individual (e.g., *my aunt*). Likewise, a pronoun like *she* (singular) cannot refer to a group expressed by a plural noun phrase (*my aunts*).

---

[3]It is also worth noting the difference between this morphological notion of number from a semantic notion of numerosity or set cardinality. Take the noun phrase '*zero teachers*', which is morphosyntactically plural (it will control plural verbal agreement) yet does not refer to a collection of multiple individuals.

However, when it comes to reference relations involving the plural pronoun *they*, the generalization is more complex. There are certain contexts where *they* and a singular noun phrase within the same sentence can have the same referent: this is known as 'singular *they*'. There are several subtypes, depending on the form and interpretation of the coreferent singular noun phrase (Konnelly et al. 2023). For instance, singular *they* can interpreted as a variable bound by a quantified noun phrase (4), paralleling bound-variable *s/he* (2b). There are also 'epicene' uses of singular *they*, where the referent is generic, ungendered, unknown, or hypothetical (5).

(4)   [Each artist]$_1$ stubbed their$_1$ toe.   **Bound-variable *they***

(5)   *Context: A car is driving recklessly, but its driver is not visible*
      Whoever [that driver]$_1$ is, they$_1$ are going to cause an accident.   **Epicene *they***

Such bound-variable and epicene usages of singular *they* have been attested since the 1400s (Balhorn 2004), and only came under critique in more recent centuries (Bodine 1975), where certain prescriptive norms have deemed the 'correct' pronoun to use in these contexts to be *he* (or perhaps a phrase like *he or she*). Despite such prescriptive norms — and even among individuals who value them — contemporary speakers of English routinely use, accept, and easily comprehend *they* in these contexts, as on- and off-line studies have shown (Foertch and Gernsbacher 1997, Han and Moulton 2022).

What the quantified and epicene usages have in common is that *they* refers to individuals whose gender is unknown, or can vary across situations. There are also at least two types of 'definite singular *they*', whose referent is a single specific individual known to the speaker. One type we call 'concealed *they*'. It is used when the speaker has in mind a single specific referent and knows their gender, but does not take for granted that all discourse participants do too. Concealed *they* is often used when the referent is not present in the discourse (6a), or if anonymity is important (6b). In contexts like these, speakers may use *they* even if the referent is a person with binary gender, for whom reference with either *he* or *she* would be obligatory in other contexts.

(6)   a.   *Context: The speaker receives a text message from a cisgender man, who the interlocutor has never met.*
           [My roommate]$_1$ says that they$_1$ are locked out of the house.

      b.   *Context: The speaker is lawyer, and is referring to client whose identity must be kept confidential.*
           [My client]$_1$ says that they$_1$ have no statement to give at this time.
                                                                                      Concealed *they*

The second major type of definite singular *they* is generally reserved for certain people with transgender or nonbinary gender identities, those who request/expect reference with *they* in all linguistic contexts, as a gender-neutral alternative to *he* or *she*. This is 'nonbinary *they*' (Konnelly et al. 2023), and its use is not restricted to concealing or backgrounding contexts; it freely corefers with a proper name, for instance (7).

(7)   *Context: Masha Gessen is a nonbinary journalist, whose pronouns of reference are they/them*
      [Masha Gessen]$_1$ said that they$_1$ will attend the party.   **Nonbinary *they***

Research in intersecting fields — sociolinguistics, formal morphosyntax, experimental syntax and semantics — identifies variation in the usage of singular *they* that is hypothesized to reflect stages of ongoing language change (Bjorkman 2017, Conrod 2019, 2022, Conrod et al. 2022, Konnelly and Cowper 2020). In Conrod's (2022) three-way model of the change, a "conservative" dialect only allows *they* in bound variable contexts, an "innovative" dialect allows *they* with any referent type (including definite specific referents, like named nonbinary individuals), and an "intermediate" dialect allows *they* with any referent that is not definite/specific, including generic or epicene ones.

Empirical evidence for dialect groupings like these comes from recent off-line acceptability studies. For example, Camilliere et al. 2021 tested the acceptability of *they* in combination with a wide range of antecedents: plural nouns, quantified singulars, definite singulars, proper names, etc. Analyzing the singular cases, the naturalness ratings generally decreased as the antecedent became 'more definite', corresponding to the more innovative uses of singular *they* (6, 7). Camilliere et al. also found several significant effects of individual differences, derived from a post-experiment survey on demographics and attitudes. Ratings of singular *they* correlated negatively with participant age, and positively with scores of gender-identity familiarity and nonbinary acceptance. The authors conducted a clustering analysis on their rating data, finding evidence for the three dialect groups described above.

Work in variationist sociolinguistics has more thoroughly investigated the social variables that predict which dialect an individual belongs to. In a large internet-based study, Conrod (2019) finds main effects of age, gender, and transgender identity on the acceptability of definite singular *they*, with younger and transgender participants rating it better. Ratings were negatively correlated with age among cisgender people but not trans people, and among men and women but not people who identify as neither. Adopting the widely accepted Apparent Time Hypothesis (Weinreich et al. 1968), the effects of age on acceptability show that the emergence of definite singular *they* is a change in progress: younger speakers' linguistic experience has a relatively higher proportion of definite singular observations of *they* than that of older speakers, hence the negative correlation. Conrod also found that the effect of age was stronger when *they* had a proper name antecedent than other definite noun phrases, suggesting that nonbinary *they* (7) is relatively more innovative than concealed *they* (6).

Further evidence that social attitudes predict acceptability of definite singular *they* come from off-line acceptability tasks and sentiment-analysis studies (Bradley 2020, Hekanaho 2020, 2022, Schultz 2021): individuals with more prescriptive views on language and with more conservative social beliefs about gender are less likely to accept innovative uses of *they*.

Note that the majority of this research has used off-line measures, like acceptability ratings. The emergence of singular *they*, especially the definite singular usages, has been characterized as a 'change from above' (Konnelly and Cowper 2020): speakers are consciously aware of innovative usages, and might adopt or reject them in different contexts depending on their social goals (cf. Labov 1966). Less well understood is the latent, unconscious status of singular *they* in the minds of individual speakers, and the extent to which off-line measures of acceptability reflect that.

2.2 Processing cataphors in real time

A cataphoric pronoun is one that precedes the nominal supplying its reference; it contrasts with an anaphoric pronoun, whose referent comes before it. One place a cataphor can appear is within in

preposed subordinate clauses, as in (8), where its nearest grammatical coreferent — its 'postcedent' — is the main-clause subject (Carden 1982, Reinhart 1983). Due to the gender features of *he* and *she* in English, a referential dependency will be infelicitous if the pronoun and its potential postcedent are mismatched in gender. Compare the following examples, where '#' marks infelicity relative to normative gender expectations.

(8) a. After she$_1$ smiled, the bride$_1$ started laughing.

b. #After she$_1$ smiled, the groom$_1$ started laughing.

c. #After he$_1$ smiled, the bride$_1$ started laughing.

d. After he$_1$ smiled, the groom$_1$ started laughing.

Comprehenders within and beyond English begin an active search for a referent for the cataphor when it is encountered, anticipating a postcedent with appropriate features at the nearest grammatical position. Seminal evidence for this active search for a postcedent comes from Van Gompel and Liversedge's (2003) eyetracking study on British English. Across sentences similar to those in (8), processing difficulty emerged just after main-clause subject nouns that were gender-incongruent with the cataphoric pronoun. For instance, the *she...*MASC (8b) condition was harder to process than the *she...*FEM (8a) condition, starting at the main clause subject region. This is known as a *Gender Mismatch Effect*. Similar effects have been observed in a range of languages and linguistic structures, demonstrating that the active processing strategy for cataphora is persistent (Giskes and Kush 2021); it is sensitive to syntactic constraints on coreference (Kazanina et al. 2007, Kush and Dillon 2021); it involves abstract grammatical predictions (Giskes and Kush 2022); and it can outweigh other expectations, e.g. for verbal subcategorization frames (Ackerman 2015).

In principle, any grammatical property of a pronoun constraining its reference possibilities should influence the search for a postcedent, not just gender. So, since *she* is a singular pronoun, in cataphoric position it should also lead the comprehender to expect a postcedent that is singular: a plural subject like *brides* should cause processing difficulty after cataphoric *she* (9b), just like a masculine subject does (8b). Likewise — ignoring for the moment the possibility of singular *they* — singular subjects should cause processing difficulty after a plural cataphor (9c vs. d)

(9) a. After she$_1$ smiled, the bride$_1$ started laughing.

b. #After she$_1$ smiled, the brides$_1$ started laughing.

c. #After they$_1$ smiled, the bride$_1$ started laughing.

d. After they$_1$ smiled, the brides$_1$ started laughing.

And indeed there is some evidence for such *Number Mismatch Effects* in cataphoric processing, paralleling the better documented Gender Mismatch Effects. In fact, one of Van Gompel and Liversedge's (2003) experiments manipulated number. They found mismatch effects at subject main-clause subjects in both *s/he...*PL (9b) and *they...*SG (9c) conditions. More recently, Giskes and Kush (2022) found similar Number Mismatch Effects in Dutch. Their study is notable because of its design, which leveraged a quirk of word order in that language. Dutch verbs agree with their subjects in number (as in English; 3), but there are certain syntactic contexts where the verb must precede the subject. Giskes and Kush found that verbal number-agreement morphology on its own

was sufficient to cause a mismatch effect after a cataphor, demonstrating that the comprehender's search for a postcedent is for abstract morphosyntactic features rather than particular nouns or nominal morphology.

Number Mismatch Effects, though, deserve deeper investigation. For reasons that might be language-general, the semantics of number is quite different from that of gender (e.g., Sudo 2012, Sauerland et al. 2005, Harbour 2014). For instance, plurals at least offer the possibility of split antecedence (10b,c).

(10) a. After they$_{1+2}$ smiled, the newlyweds$_{1+2}$ started laughing.

b. ?After she$_1$ smiled, the newlyweds$_{1+2}$ started laughing.

c. ?After they$_{1+2}$ smiled, the bride$_1$ started laughing.

There are also standard usages of *they* where the pronoun refers to a generic group of people or institution (Kitagawa and Lehrer 1990).

(11) a. They say that love is blind.

b. They make great wine in Italy.   **Generic/Institutional *they***

We are not aware of sentence-processing research which attempts to disentangle referential plural, bound-variable, and generic/institutional uses of *they* in cataphora. However, there has been some work on bound-variable and epicene *they* in anaphora. Foertch and Gernsbacher (1997) analyzed reading times of sentences with pronouns whose antecedents were indefinite or definite noun phrases, with or without lexical gender biases. In their study, conditions with *they* only had a processing advantage when the antecedent was an indefinite pronoun like *anybody*. In an eyetracking experiment, Sanford and Filik (2007) found anaphoric Number Mismatch Effects for *s/he* and *they* preceded by plural or singular indefinite antecedents, respectively. However, the effect was asymmetrical: the PL...*s/he* mismatch impeded processing in early and late eye-movement measures; the SG...*they* effect manifested only in late measures.

More recently, Han, Moulton, and colleagues have investigated both off-line acceptability and real-time processing of anaphoric singular *they*. Moulton et al. (2020) found degraded acceptability for *they* in contexts with an established gender-neutral referent that is not mentioned sentence-internally. Han and Moulton (2022) compared quantified and referential antecedents with or without nouns with gender bias. They found an acceptability advantage for bound-variable *they* relative to *s/he* when the restrictor was gender-neutral, and a disadvantage for *they* with either kind of antecedent/binder. In corresponding self-paced reading experiments, they found longer reading times when *they* had a referential, gender-biased antecedent. Moulton et al. (2022) also found high acceptability for bound-variable *they*, especially when quantifier was *each*; in their self-paced reading studies, bound-variable *they* generally facilitated processing relative to *s/he*, whatever the quantified noun's gender bias.

2.3 Hypotheses

Previous sociolinguistic work finds that age and gender identity are good predictors of the acceptability of singular *they*; previous psycholinguistic work finds that potential cataphors evoke an

active search for feature-matched postcedents. Connecting these literatures, we hypothesize that younger and transgender/nonbinary comprehenders are more likely to posit singular postcedents to cataphoric *they* during real-time sentence comprehension than are older and cisgender comprehenders. All social groups, though, should have strong expectations for singular postcedents to cataphoric *he* and *she*, since the usage of those pronouns is not undergoing significant language change. These real-time expectations will be reflected as Number Mismatch Effects: longer reading times at or just after a potential postcedent (main-clause subject) with number features discordant with the preceding cataphor, relative to concordant postcedents. The following table summarizes our predictions.

	More innovative		Less innovative	
	sg postcedent	pl postcedent	sg postcedent	pl postcedent
cataphoric *s/he*	no NME	**NME**	no NME	**NME**
cataphoric *they*	**no NME**	no NME	**NME**	no NME

Table 1: Predicted distribution of Number Mismatch Effects across comprehenders more or less innovative with respect to usage of singular *they*

A few assumptions and linking hypotheses should be noted. First, we assume that off-line acceptability judgements and real-time comprehension behavior straightforwardly reflect both each other and a more abstract notion of individuals' internalized language. Especially since the emergence of singular *they* is a change from above, one that speakers are consciously aware of, this need not be the case. Individuals might inflate or deflate their acceptability judgements to conform to implicit social pressures or internalized prescriptive norms, in ways that might not match their subconscious comprehension behavior. Some speakers who rate singular *they* highly might nevertheless struggle to process it in real time, and speakers who rate it poorly might nevertheless process it very easily. Second, we take age and gender identity to be reliable proxies for pronoun innovation. The present study does not involve any direct measures for acceptability of or familiarity with innovative usages of *they*. We flag these issues for future work, theoretical and empirical, in socio-psycholinguistics.

3 Experiment

The present study tests the hypotheses laid out above about the real-time comprehension of singular and plural cataphors across social groups. This section describes the experimental design and reports reading-time results.

3.1 Design

Materials

Thirty two itemsets were constructed in which a subordinate clause, containing a potentially cataphoric pronoun, precedes a main clause with noun phrases offering potential postcedents. The main-clause subject was always a definite noun, chosen from a list of gender-normed lexical items

found to have neither a strong masculine nor feminine bias (Misersky et al. 2014). A sample itemset follows.

(12) a. *s/he*...SG
       When **she** exercises at home, **the reporter** misses the librarians' enthusiastic encouragement.

   b. *s/he*...PL
       When **she** exercises at home, **the reporters** miss the librarian's enthusiastic encouragement.

   c. *they*...SG
       When **they** exercise at home, **the reporter** misses the librarians' enthusiastic encouragement.

   d. *they*...PL
       When **they** exercise at home, **the reporters** miss the librarian's enthusiastic encouragement.

Using a 2×2 design, we manipulated the potentially cataphoric pronoun (*s/he* or *they*) and the number features of the main-clause subject (singular or plural). Note that the main-clause clause also always contained a second noun with the opposite number feature; thus some cataphoric dependency was always globally available in the mismatch conditions (12b,c), even for participants less familiar with definite singular *they*.[4]

Sixty four filler sentences of comparable length and complexity were also constructed. Target items were distributed according to the Latin Square method, and shuffled among the fillers. Half of all items were followed by a comprehension question, taking the form of a sentence-recognition probe (*Was this the sentence you just read?*). However, due to a coding error, experimental software did not log responses to the comprehension questions.

Methods and procedure

Stimuli were presented using the Lexicality Maze methodology (Freedman and Forster 1985, Boyce et al. 2020), in an internet-based experiment hosted on PCIbex (Zehr and Schwarz 2018). The L-Maze is a combination of self-paced reading and a forced-choice lexical decision task. At each point in the stimulus, participants are shown two words: a real word and a nonce word, displayed side by side in a random order.[5] Participants were instructed to select the real word using the 'e' or 'i' keys on their keyboard. A correct choice would display the next pair of words; the real word of that pair would continue the sentence; and so on. An incorrect choice would prompt a feedback message, after which the participant could try again and continue with the rest of the sentence. Figure 1 schematizes an L-Maze trial.

After consenting to the study, participants read instructions and were introduced to the L-Maze methodology with three practice trials. There were two opportunities during the experiment for participants to take a short break. Upon completion, there were a few optional debriefing questions.

Participants

---

[4]Though no norming study was conducted to assess the coherence relations and possible referential dependencies between the clauses, care was taken to make either noun a plausible postcedent to the cataphor.

[5]Nonce words were generated using orthographic trigram frequencies calculated from Mary Shelley's *Frankenstein*.
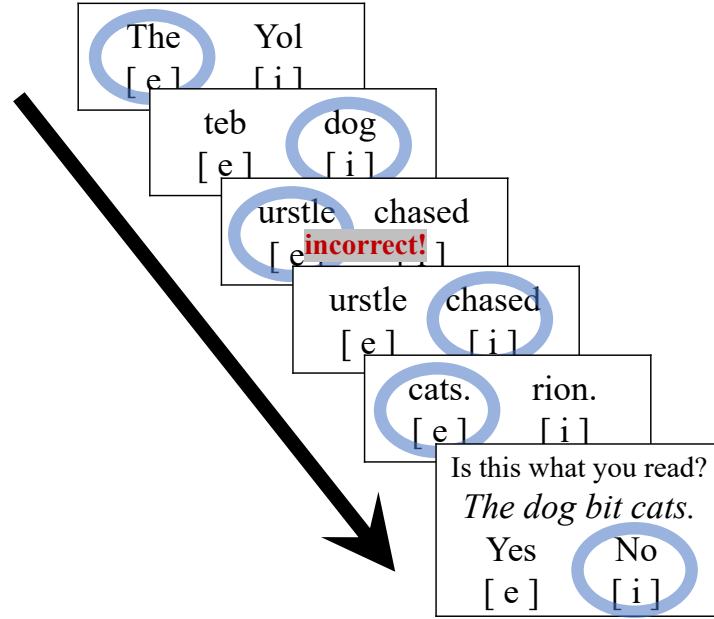
Figure 1: Illustration of the L-Maze methodology. Participant selections, input with the keyboard, are indicated with blue circles.

125 participants living in the United States were recruited via Prolific, using the platform's demographic filters to find participants of different gender identities. Five of them listed a native language other than English in their Prolific demographic profiles; we set aside their data for all analyses.

The remaining 120 participants comprised 60 transgender and nonbinary people, and 60 cisgender men and women. Age was not carefully controlled during recruitment, but the sample ended up representing a range of generations. The median age across all participants was 32 years. A more detailed breakdown is provided in Table 2.

		N$_{Subj}$	Age range (median)
Trans/nonbinary	**Trans men**	13	19–49 (31.5)
	Trans women	6	20–42 (24)
	Other gender identity	41	19–75 (32)
Cisgender	**Cis men**	38	18–65 (32.5)
	Cis women	22	21–79 (45)
All participants		120	18–79 (32)

Table 2: Breakdown of participant demographics; ages given in years

Analysis

Within any trial, observations at or after an incorrect lexicality decision were excluded from analysis. The remaining word-by-word reading times (86% of all observations) were log-transformed and analyzed with linear mixed effects models, using the R package *lme4* (Bates et al. 2015). The

fixed effects were CATAPHOR (*s/he* or *they*), MATCH (concordance in morphosyntactic number between the cataphor and main-clause subject), and either AGE or GENDER. The grammatical factors were sum-coded in the following way: *s/he* = -0.5 (12a,c) and *they* = $+0.5$ (12b,d); *number match* = -0.5 (12a,d) and *number mismatch* = $+0.5$ (12b,c). As for the demographic variables, AGE[6] was treated as a continuous factor; GENDER was sum-coded: *transgender/nonbinary* = -0.5 and *cisgender* = $+0.5$. Complexity of random-effect structure was decreased until models converged (Barr et al. 2013).

3.2   Results

Log-transformed word-by-word reaction times pooled across all participants are plotted in Figure 2; plots in Figures 3 and 4 partition RTs at critical regions by age and gender identity. Visual inspection suggests a Number Mismatch Effect emerging at the main-clause subject noun region in the *s/he*...PL condition (12b; light grey) for all participants, which persists into the spillover region (w08). Insofar as there is a parallel effect for the *they*...SG condition (12c; light gold), it does not emerge until the spillover region. Comparing across demographic bins, numerical trends align with our hypothesis: on average, it appears that older and cisgender comprehenders have more pronounced Number Mismatch Effects in the *they*...SG condition, particularly in the spillover region after the noun.
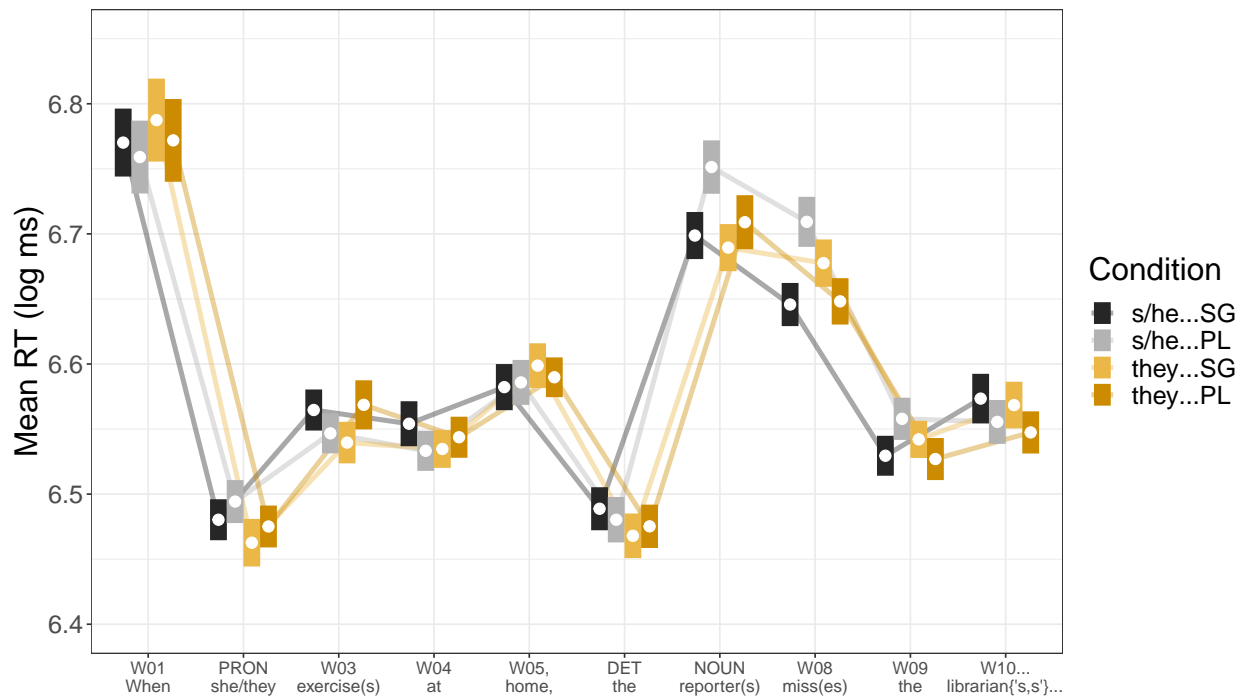

Figure 2: Word-by-word L-Maze reaction times. White dots are means; bars indicate plus/minus one standard error, calculated by participant.

---

[6]i.e., age on December 31, 2023, the year of data collection; Prolific demographic information only includes birth year. One nonbinary participant's Prolific demographics did not list their birth year; they were excluded from age-related analyses.
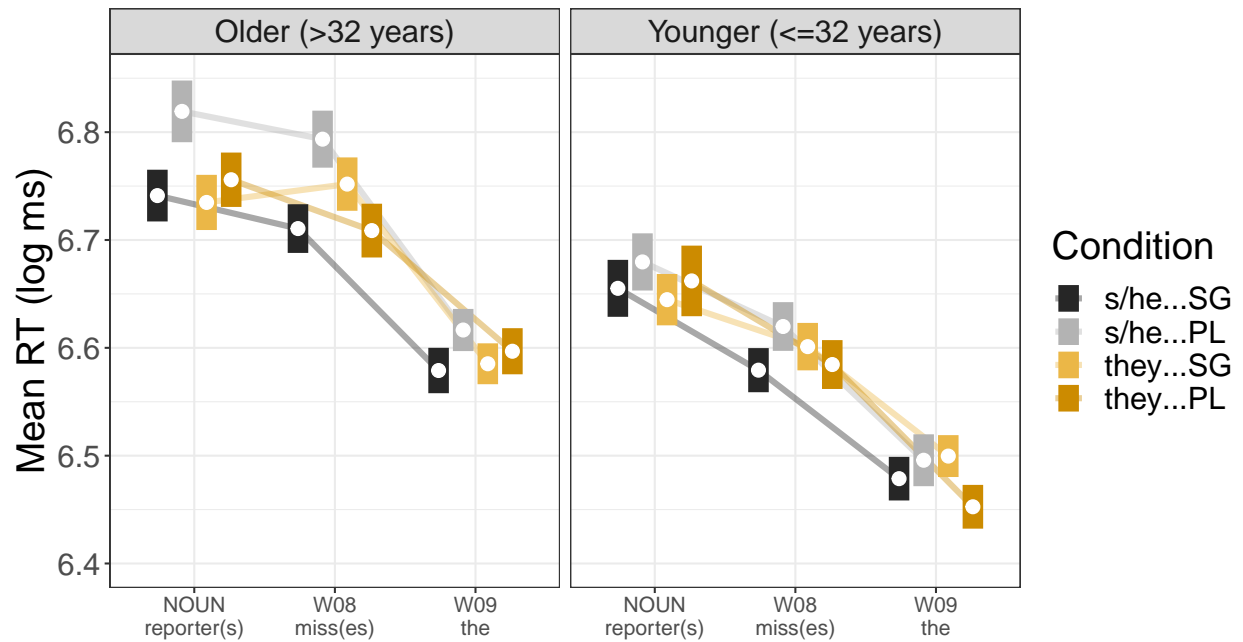
Figure 3: L-Maze reaction times at the critical main-clause subject noun and two spillover regions, partitioned by median participant age (32 years). White dots are means; bars indicate plus/minus one standard error, calculated by participant.
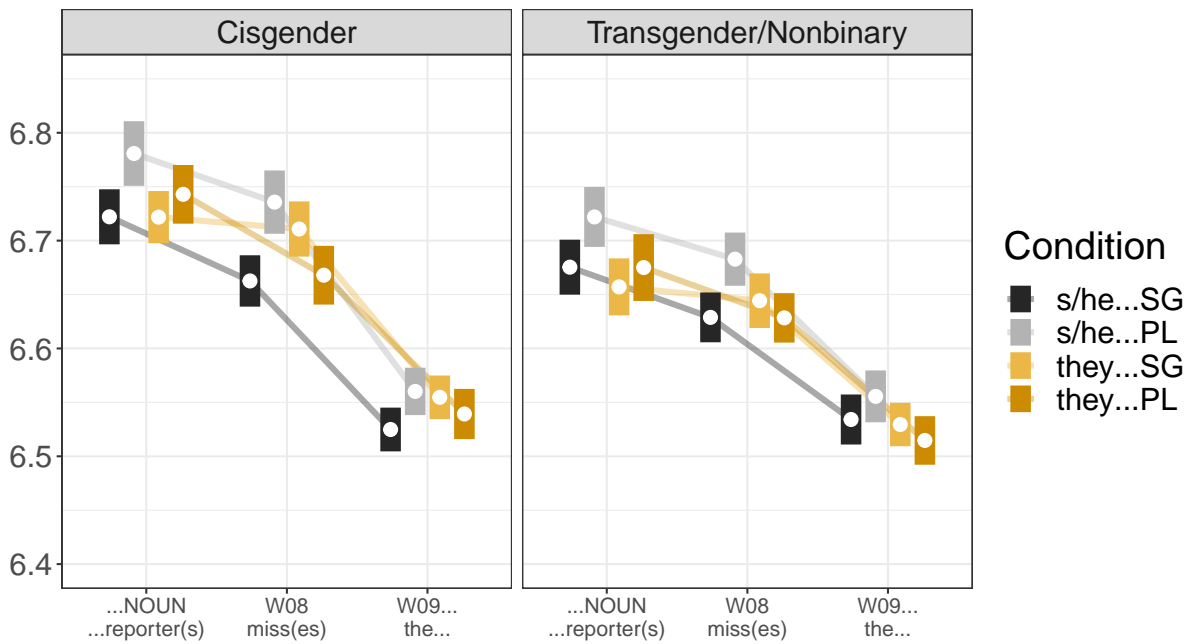

Figure 4: L-Maze reaction times at the critical main-clause subject noun and two spillover regions, partitioned by gender identity. White dots are means; bars indicate plus/minus one standard error, calculated by participant.

Results of linear models are reported in the following tables. First, consider the effects of the grammatical manipulations, without age or gender as a predictor (Table 3). At the postcedent noun region, there is a significant main effect of Cataphor (on average, conditions with *s/he* were read more slowly than those with *they*), a marginal main effect of match (mismatched conditions seemingly slower, on average), and a significant Cataphor:Match interaction (the mismatch effect is stronger for *s/he* conditions). In the first spillover region after the main-clause noun, the main effect of Match reaches significance, and the Cataphor:Match interaction persists. In the second spillover region, the main effect of Match persists.

Noun region

LogRT ∼ Cataphor*Match + (1+Cataphor+Match|Participant) + (1|Item)

	Est.	SE	df	t	p	
(Intercept)	6.7	0.025	85	260	< 0.001	***
Cataphor	-0.023	0.0093	2900	-2.5	0.011	*
Match	0.018	0.0099	110	1.8	0.063	.
Cataphor:Match	-0.069	0.018	3200	-3.6	< 0.001	***

First spillover region

LogRT ∼ Cataphor*Match + (1+Cataphor|Participant) + (1+Match|Itemset)

	Est.	SE	df	t	p	
(Intercept)	6.6	0.026	63	250	< 0.001	***
Cataphor	-0.014	0.0096	110	-1.5	0.12	
Match	0.046	0.010	28	4.4	< 0.001	***
Cataphor:Match	-0.042	0.018	3100	-2.2	0.025	*

Second spillover region

LogRT ∼ Cataphor*Match + (1+Cataphor+Match|Participant) + (1+Match|Item)

	Est.	SE	df	t	p	
(Intercept)	6.5	0.022	61	280	< 0.001	***
Cataphor	-0.012	0.0089	110	-1.3	0.18	
Match	0.024	0.0096	27	2.5	0.017	*
Cataphor:Match	-0.0078	0.016	3000	-0.47	0.63	

Table 3: Results of linear mixed-effect modeling of log RTs at the main-clause subject noun region and two spillover regions, without demographic fixed effects. Random-effects structure given in *lmer* syntax.

Next, consider the models using participant age as a continuous fixed effect (Table 4; p. 20). At all three analyzed regions, there is a significant main effect of Age; RTs reliably increase with age, as a baseline. At the subject noun region, there is a significant Match:Age interaction (the size of the Number Mismatch Effect increases with age) and a significant Cataphor:Match:Age

interaction (the *s/he*...ᴘʟ effect increases with age). At the first spillover region, the three-way interaction is merely marginally significant. At the second spillover region, there is a significant Cᴀᴛᴀᴘʜᴏʀ:Mᴀᴛᴄʜ interaction (on average, the Number Mismatch effect is greater given *s/he*) and a significant Cᴀᴛᴀᴘʜᴏʀ:Mᴀᴛᴄʜ:Aɢᴇ interaction (as age increases, the *they*...ꜱɢ mismatch effect decreases relative to the *she*...ᴘʟ effect).

Finally, the models with gender identity (cis vs. non-cis) as a fixed effect (Table 5; p. 21). At the main-clause noun, there is a significant main effect of Cᴀᴛᴀᴘʜᴏʀ (on average, conditions with *s/he* were read more slowly), marginal main effects of Mᴀᴛᴄʜ (on average, mismatched conditions seem slower) and Gᴇɴᴅᴇʀ (on average, cisgender participants seemed to read more slowly), and a significant Cᴀᴛᴀᴘʜᴏʀ:Mᴀᴛᴄʜ interaction (the *s/he*...ᴘʟ effect being greater). At the spillover region, the main effect of Mᴀᴛᴄʜ reaches significance, and the Cᴀᴛᴀᴘʜᴏʀ:Mᴀᴛᴄʜ interaction persists. At the second spillover region, the main effect of Mᴀᴛᴄʜ persists, and there is marginally significant Cᴀᴛᴀᴘʜᴏʀ:Gᴇɴᴅᴇʀ interaction (at this region, cisgender participants read conditions with cataphoric *they* more slowly).

4   Discussion

Here we provide interpretations for the experiment's results, occasionally speculating to lay groundwork for future research. One clear finding is evidence for Number Mismatch Effects in the processing of cataphora. However, the effect was asymmetrical: a *s/he*...ᴘʟ mismatch was generally harder to process than a *they*...ꜱɢ mismatch. Comparing reading times at the critical number-inflected noun region and the following spillover words, there is evidence of a temporal asymmetry too: the *they*...ꜱɢ mismatch generally emerges later, only in spillover regions. So, upon encountering a singular pronoun in a potentially cataphoric position, comprehenders form a strong expectation for a singular postcedent in the closest grammatical position. The incompatibility of a plural noun is recognized immediately, and the processing cost of this foiled expectation lingers into the next word region. But, upon encountering cataphoric *they*, the expectations for a plural postcedent seem to be weaker, or at least take more time to verify.

This asymmetry is consistent with some previous studies on anaphora. Mismatches between morphosyntactic number of an antecedent noun and subsequent pronoun are registered with a delay, and elicit smaller processing costs in ꜱɢ...*they* configurations (Sanford and Filik 2007, Filik et al. 2008). Perhaps this reflects the possibility of split antecedence (10) or generic uses of *they* (11). Those usages mean that *they* is incrementally compatible with a wider range of interpretations than *s/he* is, for all English speakers — even the least innovative ones, whose use singular *they* only in very restricted contexts. In other words, *they* will always offer more ways of recovering from what could be a number mismatch at the site of the main-clause subject,[7] possibly explaining our asymmetrical findings as a kind of ambiguity advantage (Traxler et al. 1998, et seq).

It may also be that the formal representation of number contributes to the processing difference. It has been argued that plurals are featurally and/or semantically underspecified (Sauerland et al. 2005, Sauerland 2008) — i.e., singulars have a [ꜱɢ] number feature, while plurals are not specified for any number feature. Under this analysis, the broad compatibility of *they* (plural) with a wide range of referents follows from that pronoun's lack of number features. This would mean

---

[7]As a reminder: in our stimuli, the first grammatically possible referent for a cataphoric pronoun is the subject of the main clause.

that cataphoric *s/he*, but not cataphoric *they*, would evoke an expectation for a main-clause subject specified for a number feature. The underspecification hypothesis would also help explain other phenomena in the processing of plural expressions — like response times in picture-matching tasks manipulating the numerosity of depicted objects referred to by definite plural noun phrases (Patson et al. 2013). But, as Patson (2014) reviews, still other findings are more compatible with a fully specified representation of plurals (i.e. [PL]), perhaps operating at different levels of linguistic/discourse representation.

In any case, it is notable that previous studies on cataphora do not find such clearly asymmetrical Number Mismatch Effects. In Van Gompel & Liversedge's (2003, exp. 3) eyetracking experiment on English, the *they*...SG condition exhibited processing disruptions similar in timecourse and magnitude to the *s/he*...PL condition. Likewise, Giskes & Kush's (2022) self-paced reading experiment on Dutch finds roughly symmetrical Number Mismatch Effects — though processing difficulty in their *they*...SG condition dissipates one region sooner than in the *s/he*...PL condition.

Why does the *they*...SG effect in the present study seem to be different? If the asymmetry simply reflects a semantic or morphosyntactic difference between singular and plural categories, all else equal, we would expect it to manifest across languages and historical moments.[8] One explanation has to do with the task: perhaps the L-Maze methodology localizes certain processing difficulties in unique ways; indeed, Maze variants have been shown to reduce spillover effects, compared to traditional self-paced reading techniques (Boyce et al. 2020). Another possible explanation is the participant populations and their languages. Van Gompel & Liversedge recruited British undergraduates in the early 2000s; Giskes & Kush recruited native speakers of Dutch. We speculate that the state of contemporary North American English specifically — exhibiting language change which is politically and culturally salient, even to speakers with non-innovative pronoun use — may be reflected in our participants' processing behavior: perhaps a baseline awareness of definite singular *they* has influenced the kinds of postcedents they might expect for potentially cataphoric *they*.

As for our demographic manipulations, sociolinguistic work (Conrod 2022, Konnelly et al. 2023) informed two predictions. Regarding age, younger participants would have relatively weaker plural expectations given cataphoric *they* than older speakers, since the emergence of definite singular *they* is a change in progress. Regarding gender identity, transgender/nonbinary participants would have weaker plural expectations given *they* than cisgender participants, since non-cis people are more likely to use nonbinary *they* (7).

It is clear that age is a good predictor of baseline reading times: on average, reaction times to the L-Maze task increased with participant age. There is also evidence, from MATCH:AGE and CATAPHOR:MATCH:AGE interactions, that older participants experience greater Number Mismatch Effects, which are more asymmetrical between the *s/he*...PL and *they*...PL cases. We offer two interpretations for these interactions between linguistic factors and age. First, it could be that comprehender age is indeed a reliable predictor of pronoun-dialect, and thus for the kinds of expectations evokes by cataphors in real time. The more dramatic *they*...SG effect among older participants reflects a stronger expectation for plural; that aligns with off-line acceptability of definite singular *they*, which tends to decrease with raters' age. An alternative interpretation is that age does not reliably predict pronoun-dialect or number-expectation; instead, the magnitude of reading-time effect size is simply proportional to baseline reading times. Since older participants tend to read more

---

[8]Though it may turn out that the semantic or morphosyntactic values associated with number categories are in fact subject to crosslinguistic variation, or historical change.

slowly, their mismatch effects and effect-asymmetries will be predictably more dramatic. It could be that *they* in fact evoked weak plural expectations among all participants, but for younger ones it was harder to detect, given their faster RT baseline.

For future work on age-effects on the processing of singular *they*, we note an important difference between these two interpretations. The first is a sociolinguistic explanation, about more-or-less coherent dialect clusters. In general, linguistic variables do not undergo change uniformly as time passes; language change generally follows 'S-shaped' curves, with generational tipping points (Weinreich et al. 1968, Tagliamonte and D'Arcy 2009). So, the magnitude of the *they*...SG mismatch effect qua sociolinguistic variable should not correlate linearly with age; rather, it should be bimodally (or polymodally) distributed, into clusters corresponding to more or less innovative dialects.

The second explanation is a lower-level psychological one: older participants read slower on average, and consequently effect sizes for them will be proportionately larger and easier to detect. If this is a better theory of age-related individual variation in this experimental paradigm, then we expect the Number Mismatch Effect to indeed increase linearly with age (or, more directly, with average baseline reading time).

As for gender identity, we found little conclusive evidence that cisgender participants process cataphors differently from transgender/nonbinary participants. This suggests that the cis/non-cis distinction is at best a noisy predictor of this point of individual linguistic variation with respect processing *they* pronouns. This is not so surprising, as our social variables are at best proxies of the relevant linguistic variable — namely, the categorical state of the comprehender's grammar, which might be conservative, intermediate, or innovative with respect to singular *they*. The state of an individual's grammar is not easy to infer, so we have used "familiarity with" or "usage of" definite specific *they* as a proxy for that, and we have treated age and gender as proxies of familiarity.

It is not hard to imagine how gender might be only indirectly related to grammatical innovation. Perhaps because some of our cisgender participants were highly innovative pronoun users, highly conscientious of gender-neutral language; perhaps some of our trans/nonbinary participants had less exposure to and practice with singular *they* than we assumed they did, and are consequently less fluent users of definite singular *they*, due to limited exposure to and practice with it. Future work might untangle these possibilities by investing more fine-grained variables (including attitudes about gender and language), more direct measures of grammatical state (e.g., categorization based on acceptability judgments from the same individuals who participate in reading tasks), or by integrating traditional variationist methodologies like sociolinguistic interviews.

5   Conclusion

A Lexicality Maze experiment, with participants representing a range of ages and gender identities, investigated links between sociolinguistic variation in the usage of singular *they* and the real-time processing of cataphors. We find evidence that all participants have strong expectations that cataphoric *s/he* will be postceded by a singular noun phrase. In contrast, expectations that *they* will have a plural postcedent noun is weaker, and especially weak among younger participants. This asymmetry tracks with acceptability studies on singular *they*, and suggests that younger comprehenders are more likely to expect innovative singular usages of the pronoun *they* during real-time sentence processing processing. Participants' gender identity (cisgender vs. non-cisgender), on the

other hand, seems to be too coarse-grained a social variable to reliably predict variation in number-expectations for cataphoric *they*.

Methodologically, our study is innovative in a few key ways. Experimental work on singular *they* has increased in recent years, but it has mostly focused on anaphora. By investigating cataphora instead, this study gives comprehenders time to dwell on possible interpretations of a pronoun, before having to evaluate them relative to the morphosyntactic and lexical-semantic properties of a coreferential nominal that supplies its referent. Cataphoric dependencies involving overt pronouns are also never obligatory in English, so the stimuli here pose less of a chance of exposing non-innovative users to referential dependencies which necessitate definite singular *they*.

Second, our study recruited equal numbers of cisgender and trans/nonbinary participants. Whether innovative pronoun use and comprehension is connected more to social attitudes (individuals' political views or epistemology of gender, say; Conrod 2022) or to acquisition (the number of tokens of definite singular *they* have heard and produced) on average it is bound to be those people with transgender, nonbinary, and other non-cisgender identities who are the most innovative. Therefore it behooves investigation of this particular domain of individual variation to actively recruit trans/nonbinary participants.

Finally, we note an important linguistic factor to investigate in future research. This study's design conflates the concealed (6) and nonbinary (7) definite-singular uses of *they*. Main-clause subject nouns were definite, but stimuli were presented out of the blue, without contexts that might support one or the other reading of singular *they*. And insofar as our participants had meaningful intuitions about what kinds of nouns (*reporter*, *senator*, *barista*, etc.) are likely to refer to nonbinary people, we did not attempt to control this when constructing stimuli. Previous work on singular *they* has noted higher acceptability when antecedents are definite nouns rather than names (Bjorkman 2017, Conrod 2019, Camilliere et al. 2021): evidence that concealed *they* is less innovative than nonbinary *they*. After all, using someone's name is generally incompatible with concealing or anonymizing their gender; a name that antecedes singular *they* is very likely to refer to a nonbinary person. How different types of definite singular *they* are identified and comprehended is an important question, one we leave to future socio-psycholinguistic work.

References

Ackerman, Lauren. 2015. Influences on parsing ambiguity. Ph.D. Dissertation, Northwestern University.

Balhorn, Mark. 2004. The rise of epicene *they*. *Journal of English Linguistics* 32:79–104.

Barr, Dale, Roger Levy, Christoph Sheepers and Harry Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68:255–278.

Bates, Douglas, Martin Mächler, Ben Bolker and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67:1–48.

Bjorkman, Bronwyn. 2017. Singular *they* and the syntactic representation of gender in English. *Glossa: A Journal of General Linguistics* 2:80.

Bodine, Ann. 1975. Androcentrism in prescriptive grammar: Singular 'They', sex-indefinite 'He', and 'He or she'. *Language in Society* 4:129–146.

Boyce, Veronica, Richard Futrell and Roger Levy. 2020. Maze made easy: Better and easier measurement of incremental processing difficulty. *Journal of Memory and Language* 111:104082.

Bradley, Evan. 2020. The influence of linguistic and social attitudes on grammaticality judgments of singular *they*. *Language Sciences* 78.

Camilliere, Sadie, Amanda Izes, Olivia Leventhal and Daniel J. Grodner. 2021. *They* is changing: Pragmatic and grammatical factors that license singular *they*. *Proceedings of the Annual Meeting of the Cognitive Science Society* 43:1542–1548.

Carden, Guy. 1982. Backwards anaphora in discourse context. *Journal of Linguistics* 18:361–387.

Chomsky, Noam. 1981. *Lectures on government and binding*. Providence, RI: Foris.

Conrod, Kirby. 2019. Pronouns raising and emerging. Ph.D. Dissertation, University of Washington.

Conrod, Kirby. 2022. Variation in English gendered pronouns: Analysis and recommendations for ethics in linguistics. *Journal of Language and Sexuality* 11:141–164.

Conrod, Kirby, Ruth Schultz and Byron Ahn. 2022. How many selves for them? In *NELS 52: Proceedings of the 52nd Annual Meeting of the North East Linguistics Society*, eds. Özge Bakay, Breanna Pratley, Eva Neu, and Payton Deal, volume 1, 163–172. University of Massachusetts Graduate Linguistics Students Association.

Filik, Ruth, Anthony J. Sanford and Harmut Leuthold. 2008. Processing pronouns without antecedents: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience* 20:1315–1326.

Foertch, Julie and Morton Ann Gernsbacher. 1997. In search of gender neutrality: Is singular *they* a cognitively efficient substitute for generic *he*? *Psychological Science* 8:106–111.

Freedman, Sandra and Kenneth Forster. 1985. The psychological status of overgenerated sentences. *Cognition* 19:101–131.

Giskes, Anna and Dave Kush. 2021. Processing cataphors: Active antecedent search is persistent. *Memory & Cognition* 49:1370–1386.

Giskes, Anna and Dave Kush. 2022. Abstract prediction of morphosyntactic features: Evidence from processing cataphors in Dutch. *Glossa Psycholinguistics* 1:1–30.

Han, Chung-hye and Keir Moulton. 2022. Processing bound-bariable singular *they*. *Canadian Journal of Linguistics* 67:267–301.

Harbour, Daniel. 2014. Paucity, abundance, and the theory of number. *Language* 90:158–229.

Heim, Irene. 2008. Features on bound pronouns. In *Phi theory*, eds. Daniel Harbour, David Adger, and Susana Béjar, 35–56. Oxford: Oxford University Press.

Hekanaho, Laura. 2020. Generic and nonbinary pronouns: Usage, acceptability, and attitudes. Ph.D. Dissertation, University of Helsinki.

Hekanaho, Laura. 2022. A thematic analysis of attitudes towards English nonbinary pronouns. *Journal of Language and Sexuality* 11:190–216.

Kazanina, Nina, Ellen F. Lau, Moti Lieberman, Masaya Yoshida and Colin Phillips. 2007. The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language* 56:384–409.

Kitagawa, Chisato and Adrienne Lehrer. 1990. Impersonal uses of personal pronouns. *Journal of Pragmatics* 14:739–759.

Konnelly, Lex, Kirby Conrod and Evan D. Bradley. 2023. *Non-binary singular they*, 450–464. New York: Routledge, 1 edition.

Konnelly, Lex and Elizabeth Cowper. 2020. Gender diversity and morphosyntax: An account of

singular *they*. *Glossa: A journal of general linguistics* 5:40–59.

Kush, Dave and Brian Dillon. 2021. Principle B constrains the processing of cataphora: Evidence for syntactic and discourse predictions. *Journal of Memory and Language* 120:104–254.

Labov, William. 1966. *he social stratification of english in new york city*. Washington, DC: Center for Applied Linguistics.

Misersky, Julia, Pascal Gygax, Paolo Canal, Ute Gabriel, Alan Garnham, Friederike Braun, Tania Chiarini, Kjellrun Englund, Adriana Hanulikova, Anton Öttl, Jana Valdrova and Lisa Von Stock-hausen. 2014. Norms on the gender perception of role nouns in zech, English, French, German, Italian, Norwegian, and Slovak. *Behavior Research Methods* 46:841–871.

Moulton, Keir, Trevor Block, Holly Gendron, Dennis Storoshenko, Jesse Weir, Sara Williamson and Chung-hye Han. 2022. Bound variable singular *they* is underspecified: The case of *all* vs. *every*. *Frontiers in Psychology* 13:880687.

Moulton, Keir, Chung-hye Han, Trevor Block, Holly Gendron and Sander Nederveen. 2020. Singular *they* in context. *Glossa: A journal of general linguistics* 5:122.

Patson, Nikole D. 2014. The processing of plural expressions. *Language and Linguistics Compass* 8:319–329.

Patson, Nikole D., Gerret George and Tessa Warren. 2013. The conceptual representation of number. *The Quarterly Journal of Experimental Psychology* 67:1349–1365.

Reinhart, Tanya. 1983. *Anaphora and Semantic Interpretation*. Routledge.

Sanford, Anthony J. and Ruth Filik. 2007. *They* as a gender-unspecified singular pronoun: Eye tracking reveals a processing cost. *Quarterly Journal of Experimental Psychology* 60:171–178.

Sauerland, Uli. 2008. On the semantic markedness of phi-features. In *Phi-theory: Phi-features across modules and interfaces*, eds. Daniel Harbour, David Adger, and Susana Béjar, 57–82. Oxford University Press.

Sauerland, Uli, Jan Anderssen and Kazuko Yatsushiro. 2005. The plural is semantically unmarked. In *Linguistic evidence: Empirical, theoretical, and computational perspectives*, eds. Stephan Kepser and Marga Reis, 413–434. Mouton de Gruyter.

Schultz, Ruth. 2021. Usage of, opinions on, and the future of a nonbinary pronoun. Princeton University undergraduate junior paper.

Sudo, Yasutada. 2012. On the semantics of phi-features on pronouns. Ph.D. Dissertation, Massachusetts Institute of Technology.

Tagliamonte, Sali A. and Alexandra D'Arcy. 2009. Peaks beyond phonology: Adolescence, incrementation, and language change. *Language* 85:58–108.

Traxler, Matthew J., Martin J. Pickering and Charles Jr. Clifton. 1998. Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language* 39:558–592.

Van Gompel, Roger P. G. and Simon P. Liversedge. 2003. The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29:128.

Weinreich, Uriel, William Labov and Marvin Herzog. 1968. *Empirical foundations for a theory of language change*. University of Texas Press.

Zehr, Jeremy and Florian Schwarz. 2018. PennController for Internet Based Experiments (IBEX). https://doi.org/10.17605/OSF.IO/MD832.

Noun region

LogRT ~ Cataphor*Match*Age + (1+Match|Participant) + (1+Cataphor|Item)

	Est.	SE	df	t	p	
(Intercept)	6.5	0.050	140	120	< 0.001	***
Cataphor	0.0066	0.028	960	0.23	0.81	
Match	-0.035	0.028	110	-1.2	0.22	
Age	0.0057	0.0012	110	4.6	<0.001	***
Cataphor:Match	0.041	0.055	3100	0.74	0.45	
Cataphor:Age	-8.3×10^{-4}	7.2×10^{-4}	3200	-1.1	0.25	
Match:Age	0.0015	7.6×10^{-4}	110	2.0	0.047	*
Cataphor:Match:Age	-0.0031	0.0014	3100	-2.1	0.034	*

First spillover region

LogRT ~ Cataphor*Match*Age + (1+Cataphor|Participant) + (1+Match|Item)

	Est.	SE	df	t	p	
(Intercept)	6.4	0.043	140	140	<0.001	***
Cataphor	0.0083	0.028	110	0.29	0.77	
Match	0.030	0.028	1000	1.0	0.27	
Age	0.0071	0.0010	110	7.0	<0.001	***
Cataphor:Match	0.049	0.056	3100	0.88	0.37	
Cataphor:Age	-6.5×10^{-4}	7.5×10^{-4}	110	-0.86	0.38	
Match:Age	4.7×10^{-4}	7.3×10^{-4}	3100	0.64	0.52	
Cataphor:Match:Age	-0.0025	0.0015	3000	-1.6	0.089	.

Second spillover region

LogRT ~ Cataphor*Match*Age + (1+Cataphor+Match|Participant) + (1+Match|Item)

	Est.	SE	df	t	p	
(Intercept)	6.3	0.038	140	150	<0.001	***
Cataphor	-0.0046	0.026	110	-0.17	0.86	
Match	0.032	0.025	110	1.2	0.20	
Age	0.0054	9.0×10^{-4}	110	6.0	<0.001	***
Cataphor:Match	0.12	0.048	2900	2.4	0.013	*
Cataphor:Age	-1.8×10^{-4}	7.1×10^{-4}	110	-0.26	0.79	
Match:Age	-2.2×10^{-4}	6.5×10^{-4}	110	-0.34	0.73	
Cataphor:Match:Age	-0.0036	0.0013	290	-2.7	0.0055	**

Table 4: Results of linear mixed-effect modeling of log RTs at the main-clause subject noun region and two spillover regions, with age as a continuous fixed effect. Random-effects structure given in *lmer* syntax.

Noun region

LogRT ~ Cataphor*Match*Gender + (0+Cataphor*Match|Participant) + (1+Gender|Item)

	Est.	SE	df	t	p	
(Intercept)	6.7	0.025	84	260	<0.001	***
Cataphor	−0.023	0.0096	2700	−2.4	0.013	*
Match	0.019	0.0099	120	1.9	0.057	.
Gender	0.058	0.035	110	1.6	0.095	.
Cataphor:Match	−0.067	0.019	2400	−3.4	<0.001	***
Cataphor:Gender	0.023	0.019	3000	1.2	0.20	
Match:Gender	0.010	0.020	120	0.51	0.60	
Cataphor:Match:Gender	−0.010	0.039	240	−0.27	0.78	

First spillover region

LogRT ~ Cataphor*Match*Gender + (1+Cat+Match+Gender|Participant) + (1+Match|Item)

	Est.	SE	df	t	p	
(Intercept)	6.6	0.026	62	250	<0.001	***
Cataphor	−0.014	0.0099	200	−1.4	0.15	
Match	0.046	0.011	32	4.0	<0.001	***
Gender	0.051	0.031	110	1.6	0.10	
Cataphor:Match	−0.042	0.018	3100	−2.2	0.023	*
Cataphor:Gender	0.0041	0.020	210	0.20	0.83	
Match:Gender	0.016	0.021	140	0.76	0.44	
Cataphor:Match:Gender	−0.0020	0.038	2800	−0.053	0.95	

Second spillover region

LogRT ~ Cataphor*Match*Gender + (1+Cataphor+Match|Participant) + (1+Match|Item)

	Est.	SE	df	t	p	
(Intercept)	6.5	0.022	61	280	<0.001	***
Cataphor	−0.011	0.0089	110	−1.3	0.18	
Match	0.024	0.0095	26	2.5	0.016	*
Gender	0.011	0.026	110	0.41	0.67	
Cataphor:Match	−0.0073	0.016	3000	−0.44	0.65	
Cataphor:Gender	0.031	0.018	110	1.7	0.088	.
Match:Gender	0.024	0.016	120	1.4	0.15	
Cataphor:Match:Gender	0.0053	0.033	2700	0.16	0.87	

Table 5: Results of linear mixed-effect modeling of log RTs at the main-clause subject noun region and two spillover regions, with gender identity as a fixed effect. Random-effects structure given in *lmer* syntax.