This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.
Perception and Production of Singular
They in British English

Nadir Junco
S2124944
B172654

Master of Science by Research Linguistics

Supervisors: Professor Lauren Hall-Lew and Doctor Hannah Rohde
Date of Submission: August 2023
Word count: 19,420
Abstract

Singular *they* has been part of the English language since at least the late 14th century, when it first appears in written records (Newman, 1992). Historically used in epicene contexts, the pronoun has been expanding to more specific contexts as part of a language change currently in progress (Conrod, 2019). One of these new uses is as a singular personal pronoun, where *they* has received increased attention, due to its signaling of gender-neutrality, inclusivity and diversity. Especially in its latter function, attitudes towards singular *they* may indicate attitudes towards queer people themselves (Hernandez 2020).

While initial studies did not take gender diversity into account, more recent studies have included or focused on data from the speakers who are most impacted by the discourse around singular *they* - trans* and non-binary people who use singular *they* as their personal pronoun. Through this shift to a trans linguistics (Zimman, 2020) viewpoint, studies have confirmed a language change currently in progress that sees singular *they* expand beyond its traditional contexts (Konnelly & Cowper, 2020; Conrod, 2019).

My study aims to add to the growing, queer-centric literature on singular *they*, by analysing the impact of language-internal and –external factors on naturalness rating, reading time and oral production data. While the rating data shows similar effects as observed in previous research, the reading time data indicates a significant demographic effect which has not been reported in previous studies. Finally, the oral production experiment does not return a significant impact of auditory priming, but some novel demographic effects.

The study provides a holistic snapshot of singular *they* in British English, underscoring the importance of diverse data collection due to the differences observed across the three data types. It further highlights the benefits of a trans linguistics approach, as the highly gender-diverse sampling reveals some novel insights that have been absent from previous research. Finally, the results of the production task, in particular the lack of a short-term priming effect, indicate the importance of explicit exposure to singular *they* such as pronoun statements and correcting cases of misgendering, to encourage the use of the pronoun. If corroborated in future research, this may help inclusive and trans-affirmative language policies, offering support for LGBTQ+ people in English-speaking countries.
Acknowledgements

First and foremost, I would like to thank my supervisors Lauren Hall-Lew and Hannah Rohde for their support, guidance and enthusiasm - it has been a great experience to benefit from their expertise. Specifically, I am grateful to Hannah for her tireless advice about the intricacies of psycholinguistic study design and data analysis in R; and to Lauren for showing me the importance of sharing one’s work, not only for research purposes but also to defend Equality, Diversity and Inclusion in the wider University- and non-academic context.

I have immensely benefitted from various sources of support at the University of Edinburgh, be it in the form of workshops or consultations, as provided by the Edinburgh Centre for Data, Culture and Society as well as the PPLS Skills centre. My thanks goes out to Lucia Michielin, James Page, James Besse, Andrew McLean, Fang Jackson-Yang, Stefano Coretta and Umberto Noe.

This dissertation would not have been possible without the people who completed the online experiments, so I would like to thank each of my participants as well as the people who allowed me to advertise the study: Al from Trans Active Edinburgh, Jordan Peyrouche from Yakety Yak Language Cafe, and Katie Nicoll Baines and Jonathan MacBride from the University of Edinburgh’s Staff Pride Network.

I would also like to thank my family and my friends for their support and encouragement, both to work hard but also to step away from the desk and enjoy life. I am greatly indebted to my friends Madeleine Jordan, for her unwavering enthusiasm and moral support, and Megan Waller for facilitating my transition to part-time work, enabling me to support myself during my studies. Last but not least, I would like to thank my partner Scott Cantisani for cheering me on and helping me overcome doubts and set-backs, as well as for lending me his services as an ad-hoc guinea pig and native speaker check.
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>Antecedent condition</td>
</tr>
<tr>
<td>APA</td>
<td>American Psychological Association</td>
</tr>
<tr>
<td>APT</td>
<td>Aversive Prejudice Theory</td>
</tr>
<tr>
<td>ERP</td>
<td>Event-related potentials</td>
</tr>
<tr>
<td>FAMD</td>
<td>Factorial Analysis of Mixed Data</td>
</tr>
<tr>
<td>GQNB</td>
<td>Genderqueer/nonbinary</td>
</tr>
<tr>
<td>ms/c</td>
<td>milliseconds per character</td>
</tr>
<tr>
<td>NB</td>
<td>Non-binary</td>
</tr>
<tr>
<td>NP</td>
<td>Noun phrase</td>
</tr>
<tr>
<td>PPLS</td>
<td>Philosophy, Psychology and Language Sciences</td>
</tr>
<tr>
<td>PSA</td>
<td>Public Service Announcement</td>
</tr>
<tr>
<td>RT</td>
<td>Reading time</td>
</tr>
<tr>
<td>UK</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>US</td>
<td>United States (of America)</td>
</tr>
<tr>
<td>VIF</td>
<td>Variance Inflation Factor</td>
</tr>
</tbody>
</table>
# Table of Contents

1. Introduction .................................................................................................................. 10

2. Literature Review ......................................................................................................... 13
   2.1. History of Singular *They* – from epicene to gender-neutral and gender-inclusive .... 13
   2.2. Experimental Linguistic Research on Singular *They* ............................................ 16
       2.2.1. Rating Studies ..................................................................................................... 20
       2.2.2. Rating and Reading Time Studies ...................................................................... 21
       2.2.3. Processing Studies ............................................................................................ 23
       2.2.4. Production Studies ............................................................................................ 25
   2.3. Motivation for Own Study ........................................................................................ 27
   2.4. Predictions ................................................................................................................ 28

3. Methods .......................................................................................................................... 31
   3.1. Materials ................................................................................................................... 31
   3.2. General Set-Up ......................................................................................................... 32
   3.3. General Data Management ...................................................................................... 33
   3.4. Norming Study ......................................................................................................... 33
   3.5. Pilot Study ................................................................................................................ 34
   3.6. Main study ................................................................................................................ 35
       3.6.1. Participant Recruitment .................................................................................... 35
       3.6.2. Participant Compensation ................................................................................ 36
       3.6.3. Structure .......................................................................................................... 36

4. Analysis and Results ....................................................................................................... 41
   4.1. Participant Demographic Overview ....................................................................... 41
   4.2. Naturalness Rating Data ......................................................................................... 44
       4.2.1. Excluded Rating Data ...................................................................................... 50
   4.3. Reading Time Data .................................................................................................. 52
       4.3.1. Excluded Reading Data .................................................................................... 56
   4.4. Oral Production Data .............................................................................................. 56
       4.4.1. Excluded Production Data ................................................................................ 62

5. Discussion ....................................................................................................................... 63
   5.1. Limitations ............................................................................................................... 63
   5.2. Naturalness Rating .................................................................................................. 64
   5.3. Reading Time ........................................................................................................... 67
   5.4. Oral Production after Auditory Priming ................................................................. 69
   5.5. General Discussion ................................................................................................. 71
6. Conclusion......................................................................................................................................................... 74
7. References .......................................................................................................................................................... 76
8. Appendix .......................................................................................................................................................... 82
   8.1. Appendix A: Norming Study ....................................................................................................................... 82
   8.2. Appendix B: Naturalness Rating Task ......................................................................................................... 83
   8.3. Appendix C: Reading Time Task .................................................................................................................. 85
   8.4. Appendix D: Production Task ...................................................................................................................... 90
      8.4.1. Phase 1: Priming/Control items .............................................................................................................. 90
      8.4.2. Phase 1: Filler items ............................................................................................................................... 91
      8.4.3. Phase 2: Production ............................................................................................................................... 92
List of Tables

Table 1 Demographic Breakdown of Norming Study Participants ........................................34

Table 2 Demographic Overview of Main Study Participants ..................................................41

Table 3 Participant Breakdown by Personal Pronouns, Gender, and Trans* Identity ........43

Table 4 Rating Model Comparisons Output ........................................................................46

Table 5 Mean Naturalness Ratings by Participant’s Self-reported Interaction with People
who go by they/them........................................................................................................49

Table 6 Rating Model Comparisons Output with Excluded Participants .........................51

Table 7 Reading Time Model Comparisons Output ..............................................................53

Table 8 Rating Time Model Comparisons Output with Excluded Participants ..................56

Table 9 Production Model Comparisons Output ..................................................................57

Table 10 Production Model Comparisons Output with Excluded Participants ..................62
List of Figures

Figure 1  *Mean Naturalness Ratings by Antecedent Condition* .......................................................... 45

Figure 2  *Mean Naturalness Ratings by Participants’ Own Personal Pronoun(s) across Antecedent Conditions* ............................................................. 48

Figure 3  *Mean Naturalness Ratings across Antecedent Conditions by Participants’ Self-reported Interaction with they/them-users* ................................................ 50

Figure 4  *Mean Naturalness Ratings across Antecedent Conditions by Participants’ Age at English Language Acquisition* ................................................................. 52

Figure 5  *Mean Reading Times across Antecedent Conditions by Participants’ Own Personal Pronouns* ............................................................................................. 55

Figure 6  *Pronoun Production Ratio by Antecedent Condition* .............................................................. 58

Figure 7  *Pronoun Production Ratio by Participants’ Own Personal Pronouns* ........................................ 59

Figure 8  *Pronoun Production Ratio by Participants’ Own Personal Pronouns and Priming Group* ........................................................................................................ 60

Figure 9  *Pronoun Production Ratio by Participants’ Self-reported Frequency of Interaction with they/them-users* ................................................................................ 61
1. Introduction

The singular they pronoun has existed in the English language for centuries, but it has been subject to criticism and attacks since the rise of prescriptivism in the 18th century (Balhorn, 2004). Historically used in contexts where the gender of the referent is irrelevant or unknown (1, 2), prescriptivists criticized what they perceived as a number mismatch as they claimed attribution of they with plural number. Despite this claim, singular they remained in use in these so-called epicene (Baron, 1981) contexts, garnering further support once second-wave feminists began the fight against the formerly prescribed generic he. In recent developments, epicene singular they has not only been formally endorsed in prestigious style guides (American Psychological Association, 2020), but singular they has also been spreading into non-epicene contexts.

Bjorkman (2017) identifies two groups of language users, depending on their acceptance of singular they: conservative users, who accept only epicene antecedents (1) and (2), compared to innovative users who also accept the pronoun with specific definite antecedents (3). Further, Konnelly and Cowper (2020) identify a third super-innovative group, who accept singular they when its corresponding referent is introduced with gendered kinship terms (4) or given names (5, 6).

1. Every person should mind their own business.
2. A caring parent would not leave their child unattended.
3. My therapist has raised their hourly fee.
4. I do have a favourite uncle; they always bring me little gifts.
5. Li can’t make up their mind about what to order.
6. You don’t have to forgive Michael just because they broke their leg.

In Konnelly and Cowper (2020)’s sample, members of this super-innovative group were either of non-binary gender or had a non-binary family member. As the pronoun may index queer gender identity of the referent or queer allyship (Conrod, 2019) in these contexts, singular they has become a target in anti-LGBTQ+ and specifically anti-transgender rhetoric. Thankfully, linguistics has also undergone a shift since its early days of prescriptivism, and sociolinguistic approaches are increasingly being employed to centre the experiences of minoritised language users, as Zimman (2020) urges in their call for a trans linguistics.
Through this dissertation, I seek to add to the growing literature of trans linguistics studies of singular they. I begin with a short overview of the history of singular they, followed by a literature review to inform the rationale behind my study. While there has been an increase in trans linguistics methods in studies that use data from rating experiments, this approach is still less frequent in processing studies, as well as those that focus on production of singular they, of which there are fewer in general. Surveying the six antecedent types above, I conduct three experiments eliciting naturalness rating, reading time, and oral production data, to answer the following questions:

1. Which language-internal and language-external factors influence the perception and production of singular they? Is the current language change being driven by language user’s mental concept of gender, or is it simply a question of frequent exposure to singular they?
2. Do rating, reading and production data show the same patterns?
3. Does a trans linguistic approach to the study of singular they lead to new insights?

While findings from trans linguistics rating studies are partially replicated, novel findings emerge in the processing and production experiments. The results of these experiments are then discussed in-depth, concluding with a call for diverse data collection in terms of recruitment and demographic data elicitation, as well as overall methodology. However, I begin with a clarification of definitions and frameworks used throughout the paper.

I will be using Ackerman (2019)’s framework of gender representation in language, which distinguishes grammatical gender (morpho-syntactically encoded feature), conceptual gender (includes semantic, definitional and notional gender; used for referent classification) and biosocial gender (a person’s gender as influenced by their biological features, gender expression, gender identity as well as socialisation and cultural norms) (p.3).

Further, I will be using the term trans*, with the asterisk denoting the inclusive nature of this umbrella term as many people of non-normative genders identify with this label. However, some people that are not cisgender may not identify as trans*. I will be using the term gender-diverse in the context of this study, to include all participants that are either trans*, of a non-cis, non-normative gender (non-binary, agender) or use other pronouns than she or he (they/them or multiple different pronouns). It should be noted that while this label of gender-diversity is used for ease of analysis and comparison within this study, the
problematic nature of unifying such a diverse group of people must be acknowledged. It is my hope that the results of this study will illuminate these differences and inspire larger-scale research that may feature more fine-grained groupings.
2. Literature Review

2.1. History of Singular *They* – from epicene to gender-neutral and gender-inclusive

My literature review is divided into two parts. First, I will provide a brief history of epicene singular *they* in the English language. This is followed by a larger section reviewing recent scholarship on singular *they*, concluding with the motivation for my own study.

Epicene singular *they* has been part of the English Language for at least six centuries, with some examples of its earliest written attestation found in Chaucer’s “Canterbury Tales” (Newman, 1992) as well as in the Helsinki Corpus of Middle English (Curzan, 2003). Drawing upon Baron (1986), Curzan notes that singular *they* did not inspire much discussion until the 18th century, when prescriptivists sought to replace it in favour of the generic *he* in written language to avoid what they posited as a number-mismatch.

The emergence and persistence of epicene singular *they* may be a result of the loss of grammatical gender in English (Newman, 1992). Analysing historical text data, Newman suggests that while the existence of grammatical gender features in Old English - possibly combined with the absence of women in public discourse - favoured the generic masculine *he* for epicene co-reference, the rise of semantic gender has led to *he* invoking salient masculine gender associations which conflicts with the required underspecification of gender in epicene contexts.

It is no surprise then that the rise of feminism starting in the 1970s challenged the prescribed generic *he* as intrinsically and discriminatorily masculine. Conrod (2020) lists multiple studies finding a lack of gender-neutral reading and instead evocation of masculine referents when using *he* in gender-neutral contexts (e.g. Moulton et al., 1978). Drawing upon previous studies that found persistent use of written singular *they* in spite of prescriptivist efforts (e.g. Meyers, 1990), Newman (1992) analysed nine US-American TV interview programmes and found a preference for singular *they* in spoken language. While this falls in line with his predictions, he struggles to explain observations of singular *they* being used with referents with strong conceptual gender association (*wife*) as well as in reference to a discourse-old specific male doctor. Denoting this specific use of the pronoun as
“problematic”, Newman suggests concealment of the referent’s identity as a possible motivation.

Less than two decades later, singular *they* is found to be used across a wide range of antecedents (Paterson, 2014). In fact, Paterson finds singular *they* used with indefinite pronouns, quantified noun phrases (NPs), indefinite NPs and definite NPs, in a corpus analysis of British English newspaper and prose publications from 2005 to 2007. To summarise then, the latter quarter of the 21st century saw singular *they* expand from epicene to gender-neutral use; it could now not only signal a lack of knowledge on the language producer’s part about the biosocial gender of the referent but also an effort to avoid association of a particular gender by the interlocutor.

At present, style manuals such as the APA finally endorse the use of singular *they* (American Psychological Association, 2020). However, as the pronoun expands beyond epicene boundaries, it faces new hurdles. In comparison with its epicene usage, relatively little scholarship explores the history of singular *they* with definite and/or specific referents. While it is difficult to pinpoint the beginning of this expansion, singular *they* as a personal pronoun for gender-diverse individuals started receiving palpably more attention towards the end of the 2010s, with a 313% increase in lookups in the Merriam-Webster dictionary in 2019 which led to it being crowned Word of the Year (*They’ Is Merriam-Webster's Word of the Year 2019*, 2019).

While gender-neutral *they* faced criticism by proponents of the generic masculine, gender-diverse *they* has come under attack by anti-trans movements. A recent example of this occurred in Florida, whose governor signed a range of bills targeting trans* children and adults, including one bill prohibiting students and teachers from stating their own pronouns in schools (Farrington, 2023). While this is an extreme example of public transphobia, other cases may be harder to identify. Similar to prescriptivists positing a number-mismatch as a result of using epicene singular *they*, the same argument is also being used to discredit gender-diverse singular *they* – an argument that may mask underlying attitudes towards trans* people themselves.

In fact, both prescriptivist and anti-trans attitudes have been found to co-occur with negative attitudes towards singular *they* in a study drawing upon Aversive Prejudice Theory (APT) (Hernandez, 2020). APT is based on the concept of Aversive Racism first coined by
Kovel (1970) and describes the phenomenon of people being implicitly prejudiced against a group while at the same time explicitly stating support for the same group, making it harder for them to acknowledge and thus confront their subconscious biases (Pearson et al., 2009). Surveying over 700 people, 74.2% of which US-Americans, the study finds that while both prescriptivism and anti-trans prejudice predict negative attitudes towards singular *they*, anti-trans attitudes are a much stronger predictor for negative attitudes towards the gender-diverse use of the pronoun compared to its gender-neutral use.

In a similar vein, sexist attitudes co-occur with negative attitudes towards gender-diverse singular *they* (Bradley, 2020). Bradley’s study examines the impact of prescriptivist and sexist attitudes on grammaticality ratings of singular *they* across different referent conditions: indefinite epicene NP, definite NP, names with strong conceptual gender associations as well as gender-neutral names invented by the author, surveying a sample of 222 mostly US-American language users, 11 of which non-binary. Distinguishing between hostile sexism, which belittles women outright, (e.g. “women are less intelligent than men”) from benevolent sexism, which uses seemingly positive language to stereotype women (e.g. “women are more nurturing than men”), Bradley finds that people that are more prescriptivist or more benevolently sexist rate definite (NP and name) singular *they* significantly lower. Further, sentences containing name antecedents received significantly lower ratings than those containing NP antecedents; this effect is not observed among non-binary participants, however. Non-binary respondents not only provided overall higher ratings for sentences containing definite singular *they* as well as those containing a conceptual “gender mismatch” (e.g. a traditionally masculine name antecedent being referred to as *she*), but also returned significantly lower scores for both hostile and benevolent sexism than binary gender respondents.

Regardless of the underlying force driving negative attitudes towards singular *they*, failure to use the pronoun has a palpable impact on discourse participants. As mentioned above, using *he* instead of *they* in gender-neutral contexts leads to interpretative bias skewing masculine, which in turn leads to exclusion of other genders. Turning to the domain of personal pronouns, referring to people with a different third person singular pronoun than the one(s) they go by constitutes an act of misgendering. Misgendering has been shown to have multiple negative effects on the recipient, such as feelings of stigmatisation and increased anxiety (McLemore, 2015), causing significant psychological distress and minority stress.
(McLemore, 2018). In the 2015 study, while pronouns are not analysed, it is observed that genderqueer individuals report being misgendered more often than trans men and trans women. Further, multiple hermeneutical and epistemic injustices may be enacted when language used by and for non-binary people is not accepted by the mainstream, e.g. denial of the reality of their lived experiences and exclusion from public discourse (Carmona, 2023).

These previously mentioned studies highlight the importance of not only researching singular they, but particularly the responsibility of the researcher when working with marginalised groups such as gender-diverse people. In the following, I will review previous linguistic studies on singular they spanning the last three decades, and illustrate the richness that trans linguistics approaches have been adding to the topic.

### 2.2. Experimental Linguistic Research on Singular They

Moving on from previously described corpus studies that sought to establish the frequency of epicene singular they in documented speech, the turn of the millennium saw the application of psycholinguistic methods to the study of singular they, as well as inclusion of materials using singular they with definite antecedents. I will now discuss three psycholinguistic studies which established the basis for modern research into singular they. This is followed by a more in-depth look into the previously mentioned papers by Bjorkman (2017) and Konnelly and Cowper (2020), whose categorisation of language users into different groups based on their acceptance of singular they with different antecedent types has laid the groundwork for the most recent scholarship.

One of the earliest psycholinguistic studies on singular they found a reading delay only with binary-gendered antecedents (Foertsch & Gernsbacher, 1997). Self-paced reading of test sentences containing singular they referring to a referent denoted by definite or indefinite NPs with either masculine, feminine or neutral conceptual gender associations, as well as indefinite pronouns, was compared to control sentences containing she or he pronouns. Results showed that they only inferred a significant cost with definite antecedents of masculine or feminine gender bias. It may be important to note that, in contrast to contemporary psycholinguistic practices, test sentences were presented in the same order to each participant and the paper does not mention filler items. Further, all experiment participants were undergraduate students at a US-American university but no demographic breakdown is provided, pre-supposing that variation is solely accounted for by antecedent
type. Another pre-supposition that is somewhat problematic from a contemporary trans linguistics reading is the labelling of they as a "plural pronoun" and its use in singular contexts as “ungrammatical”. Nevertheless, Foertsch and Gernsbacher’s study marks an important turning point in singular they scholarship that inspired a new wave of interest and research.

Sanford and Filik (2007) sought to improve the aforementioned study design through implementation of eye-tracking measures instead of self-paced reading, in order to test two different accounts of pronoun resolution. They term Foertsch and Gernsbacher (1997)'s account as the “unspecified-singular-match account”, wherein they triggers a simultaneous search for plural or gender-neutral singular antecedents. This is contrasted with Sanford and Filik’s account which posits an initial search for plural antecedents which, only in case of failure, triggers a subsequent accommodation of singular antecedents of underspecified gender. Focussing on antecedents without strong conceptual gender connotations, the study featured a 2x2 design of singular vs plural antecedent with they vs he/she pronouns – unlike Foertsch and Gernsbacher, no distinction was made between different types of singular antecedent such as indefinite pronouns compared to NPs. Due to the different methods, materials and analyses, it does not come as a surprise that Sanford and Filik reported different findings, namely a significant delay of singular they compared to plural they. This is taken to support the author’s proposed account of they reference resolution. However, when comparing singular they to singular he and she, no significant reading delay was found after correcting measurements to account for difference in character count per sentence, just as in Foertsch and Gernsbacher’s study. The assumption, then, that they is a plural pronoun to begin with, may render Sanford and Filik’s argument somewhat weaker as there is no evidence per se that they triggers a search for a plural antecedent just because singular they sentences are read at a slower pace than plural they – if this was indeed the case, one would expect singular they to show significantly slower reading times than singular he and she, which do not trigger an initial search for plural antecedents.

While Sanford and Filik (2007) provide more technically fine-grained measurements, there is little progress in the underlying assumptions towards singular they despite a decade having passed since Foertsch and Gernsbacher (1997). Although singular they is no longer described as “ungrammatical”, it is assumed to cause a number mis-match, and persistently presented as a plural pronoun. Another similarity is the use of a possibly homogenous
participant sample of undergraduate students and a lack of demographic information apart from the location of the University (Scotland in Sanford and Filik’s study). However, both studies provide valuable insights into early modern research of singular they.

Another decade later, singular they is found to receive lower ratings if the antecedent evokes stronger expectations of gender, whereas this effect is not found in reading time (RT) measures (Doherty & Conklin, 2017). One of the first singular they studies combining two different data types (naturalness rating data and eye-tracking reading data), the authors posit a semantic influence of conceptual gender expectancy as the driving force behind pronoun processing and propose a new real-world-knowledge/probability-based account of co-reference resolution. To test this, participants (undergraduate students at a UK university) rate or read sentences containing singular they, he or she with antecedents of three distinct degrees of gender expectancy as determined in a norming study: gender-known, high-expectancy and low-expectancy. Results show that naturalness ratings follow the scale predictions, with low-expectancy items eliciting the highest ratings, followed by high-expectancy and then gender-known items. In accordance with Foertsch and Gernsbacher (1997), Doherty and Conklin find no reading time delay for singular they with low-expectancy antecedents compared to the other singular third person pronouns. However, unlike in the rating data, the predicted genericness scale of gender-known > high-expectancy > low-expectancy is not mirrored in reading measures as they with gender-known antecedents shows significantly faster reading times than with high-expectancy antecedents in some regions. The authors offer a possible explanation in that interlocutors may leave the gender of gender-known antecedents as unresolved if they are referred to with they. In other words – if understood correctly -, singular they is interpreted as gender-neutral, whereas with high-expectancy antecedents, it may be interpreted as either gender-neutral or epicene and the competition between these two interpretations may result in a delay.

This study then presents a step forward towards sociolinguistic analysis of singular they by integrating real-world knowledge, in this case conceptual gender expectancy. However, despite questioning whether they is, in fact, a plural pronoun, the authors still describe singular they as ungrammatical due to number-mismatch. As with the two previous studies, all participants are undergraduate students and no detailed demographic overview is provided apart from stating that participants were either male or female.
A sociolinguistic approach towards singular *they* reveals a possible language change in progress, as more innovative language users start to accept the pronoun with not just epicene antecedents (Bjorkman, 2017). Drawing upon the author’s acceptability judgements in consultation with others, as well as observation of written data, two distinct groups of language users are identified, based on their attitudes towards singular *they*: on the one hand, conservative users, who accept the pronoun with quantificational NP and indefinite, epicene antecedents, and on the other hand, innovative users, who also accept singular *they* with definite specific antecedents but not with given names or specific NPs that carry strong conceptual gender connotations. Bjorkman proposes that the difference between these two groups is caused by the nature of gender features: for conservative users, they are contrastive and singular *they* is only felicitous in cases of underspecification. For innovative users, in contrast, gender features are optional so that singular *they* is felicitous unless those features are present, for example with given names or kinship terms. Further, Bjorkman argues that bound variables (i.e. pronouns with a quantified NP antecedent), unlike referential pronouns, may lack gender features, rendering quantificational NPs as acceptable with singular *they* even for conservative users.

While Bjorkman (2017) analyses conservative and innovative language users, she acknowledges in a footnote that some language users may accept singular *they* with name antecedents. These users not only exist in theory, but in fact form a third, super-innovative group (Konnelly & Cowper, 2020). Through application of a trans linguistics framework, collecting grammaticality judgements from eight language users who are either non-binary themselves or have a non-binary family member, Konnelly and Cowper find that all respondents accept singular *they* across all antecedent conditions (ACs). Their work underlines the important connection between language users’ mental concept of biosocial gender and their grammaticality judgements, confirming findings from Hernandez (2020). Konnelly and Cowper re-frame the three distinct group of *they*-users as chronological steps of a language change currently in progress. While they agree with Bjorkman’s analysis of conservative (= stage 1) users, they argue for only a minimal difference distinguishing them from innovative users. At the innovative stage 2, gender features are still contrastive as at stage 1, but the difference lies in the assignment of features. Unlike stage 1, assigning gender features is optional at stage 2 for some lexical items. However, nouns such as kinship terms or given names carry lexical gender features and therefore their referents do not accept singular *they* co-reference. In order to move to stage 3, gender features would need to change
from contrastive to optional, a possibility which is also available at stage 1 meaning that language users may skip stage 2.

Konnelly and Cowper (2020) show the importance of respecting and supporting the communities, especially marginalised communities, at the centre of one’s research. Their work continues the trajectory of an important turning point in the study of singular *they* towards a trans linguistic approach which can be seen from 2018 onwards. In the second part of the review, I will focus on modern studies that incorporate primarily rating data, moving on to more recent psycholinguistic work and finally, experimental studies on production of singular *they*.

### 2.2.1. Rating Studies

Conrod (2019) presents one of the biggest and most impactful works on singular *they* in American English. Their thesis is built on naturalness rating data from over 700 participants, as well as a smaller sociolinguistic interview study. After Bjorkman (2017), Conrod investigates singular *they* with four different antecedent types of proposed declining naturalness: quantified NP, definite generic NP, gender-neutral names and feminine or masculine names. Testing 15 critical items, a significant main effect of antecedent type is found, with highest ratings for generic definites and lowest for feminine names. Further, their extensive demographic analysis reveals novel discoveries. Participant age is found to be a significant predictor: as older participants rate singular *they* generally lower than younger participants, especially with name antecedents, a language change in progress is posited. Further, non-binary language users rated *they* significantly higher than men and women did, as did transgender participants compared to cisgender ones. What is more, the antecedent rating pattern is reversed in the trans*/non-binary subsample, as names are rated higher than quantified NPs, demonstrating the danger of extending outcomes from cisgender samples across all communities. Conrod further notes that metalinguistic comments by participants strongly suggest that singular *they* has come to index not only queer gender identity but also allyship and progressive values, following Silverstein (2003) model of indexical order and Eckert (2008)’s indexical field.

Moving away from syntactic and towards pragmatic accounts of singular *they*, both conceptual antecedent gender and social distance/closeness significantly impact naturalness ratings of singular *they* in American English (Camilliere et al., 2021). Findings reveal that
non-gendered antecedents are rated higher than gendered ones, and social distance also correlates with higher ratings compared to closeness, leading the authors to attribute inter- and intra-participant variability to pragmatic pressures. Moreover, younger people provide higher ratings, as do people who are more familiar with, and those more accepting of, non-binary gender identities. The study also confirms three distinct rating clusters that correspond to those identified by Konnelly and Cowper (2020). However, although reading times were collected as part of the experiment, no significant findings are reported. Due to the main effect of social distance being present even in the stage 3 group, Camilliere et al. infer that singular they is still stabilising in the grammar at all stages. It must be noted, however, that only 3 out of 160 participants recorded their gender as different from female or male, with no data on trans* status collected.

### 2.2.2. Rating and Reading Time Studies

Combining analyses of naturalness rating and on-line reading data, the two data types show differing impacts of language-internal and -external factors on singular they in Canadian English (Block, 2019). Test items of said study contain indefinite antecedents and name antecedents of either high or low conceptual gender expectancy in reference with singular they, he or she, the latter two of which are distributed to pose gender “matches” and “mis-matches”. For indefinite antecedents, longer reading times are reported for high-expectancy items than low-expectancy, as well as longer reading times for she and he (regardless of whether there was a gender “match” or “mis-match”) compared to singular they items. For name antecedents, on the other hand, sentences containing singular they show significantly longer reading times than those containing he or she (no “gender mis-match” sentences were included for name antecedents). Regarding participant gender, findings do not show a significant difference between participants of binary (n = 54) compared to non-binary (n = 29) genders, nor any non-binary exposure effects within the binary participant subset. This stands in contrast with outcomes of the rating experiment, which shows significantly higher ratings for sentences containing singular they by participants that are non-binary or report frequent exposure to non-binary pronoun they. It might be noteworthy that, like Camilliere et al. (2021) and Han and Moulton (2022), Block’s design does not include filler items that do not feature singular they.

Further differences are found in a study comparing acceptability ratings and reading times of bound/quantified (7) and referential (8) singular they (Han & Moulton, 2022).
(7) Every teacher likes their students.

(8) The teacher likes their students.

With antecedents that have no strong conceptual gender connotations, participants rate bound they higher than referential they. This pattern is reversed for masculine-gendered antecedents, which incur lower ratings with they than with he, with referential they sentences being rated particularly low. While there are no differences in reading time between bound and referential they with non-gendered antecedents, both are read slower than sentences containing bound and referential he. If the antecedent is conceptually gendered, however, referential they is read significantly slower than bound they, as well as slower than both bound and referential he. Based on these results, Han and Moulton provide a new syntactic theory of singular they that distinguishes between quantified antecedents, which do not trigger “gender-matching” constraints in bound pronouns, and referential antecedents which do require “gender-matching”. Under this view, singular they resolution is only disadvantaged with “gendered” referential antecedents. The study does not include demographic participant breakdown or analysis, suggesting a uniform underlying syntax for all language users. Further, singular they sentences are compared to their he-equivalent only, assuming that she-sentences would behave accordingly – this may be problematic due to historically more epicene qualities of he.

In a hugely influential body of work, Ackerman (2018a, 2018c, 2018d) revitalises psycholinguistic analysis of singular they through the application of trans linguistics. By considering the influence of sociolinguistic variables, Ackerman enriches previous theories of pronoun processing. A key finding, increased exposure to gender-nonconforming individuals correlates not only with higher acceptability ratings, but also faster reading times with gender-biased name antecedents, suggesting that the language change of singular they may be driven by real-life experience/exposure (Ackerman, 2018a, 2018d). Based on this inextricable link between gender non-conformity and singular they, Ackerman (2018b) highlights the care that researchers need to apply in their work on the pronoun, as previous work had failed to do so. They argue that describing singular they as “ungrammatical” and “unacceptable” is not only factually incorrect but may directly harm and delegitimise communities and individuals who use the pronoun. Further, Ackerman emphasizes the importance of questioning assumptions deep-rooted in the discipline of linguistics; in addition to the traditional misclassification of they as a plural pronoun, they question the
conclusion that reading delays automatically hint at processing difficulties or ungrammaticality. Comparing generic and specific antecedent with and without conceptual gender bias, longer reading times and more instances of re-reading occur with unbiased antecedents, as well as with specific (name) antecedents (Ackerman, 2018c). Ackerman suggests that, apart from processing difficulties, another possible cause may be participants re-adjusting their mental representation of the referent’s gender.

2.2.3. Processing Studies

Possible support for this explanation may be found in the study of event-related potentials (ERPs), which measure electrical activity of the brain in response to stimuli. However, studies using these ERP measures to test singular they processing do report measures indicating processing difficulties (Chen et al., 2021; Prasad & Morris, 2020).

Drawing from Ackerman (2019)’s three-tiered model of gender encoding, Prasad and Morris (2020) hypothesise that frequent real-world exposure to non-binary individuals will correlate with an absence of processing difficulties, which manifest as so-called P600s. ERP measures while listening to materials containing singular they are collected across four different antecedent conditions: non-referential ambiguous (indefinite pronouns), non-referential unambiguous (quantified NP with binary conceptual gender bias), referential ambiguous (definite NP without conceptual gender bias) and referential unambiguous (given names). Based on Bjorkman (2017), “super-creative” (all of which US-American) users of singular they are recruited to test whether they show no processing difficulties for singular they with name antecedents. The authors report that this hypothesis is not borne out, as they find a significantly greater P600 effect for given names compared with all other antecedent conditions.

However, some issues emerge in Prasad and Morris (2020)’s approach, first and foremost their calculation of P600 measures for singular they. Instead of elicitation through comparison with singular he or she, which were collected in another part of the experiment, singular they was compared to a baseline condition of plural they, effectively transforming the measurement into a comparison between plural and singular processing rather than singular they compared to other third person singular pronouns. I would speculate that, in this case, a P600 comparison to singular and plural you would have provided more insight. Further, test sentences contained the reflexive themselves, which has been shown to be in
competition with its variant *themself* (Ackerman, 2018a), which is not discussed in the paper. Finally, an issue that is acknowledged by the authors, is the sampling used to select stage 3 users of singular *they*. Participants were selected based on a questionnaire about their own gender identity, as well as familiarity with gender-neutral pronouns and frequent interaction with non-binary people. Prasad and Morris report through another measure, N400 (frontal negativity), that participants did not seem to dissociate binary gender features from given names. This may have been avoided if participants had been selected through a rating survey of singular *they* instead.

A P600 effect is also reported when reading materials containing singular *they* (Chen et al., 2021). This study shows a partial replication of the aforementioned issue of categorising singular *they* as a plural pronoun, and thus eliciting P600 measures through a comparison with plural *they*. Analysing reading data from undergraduate students at a progressive US-American university that teaches students about non-binary pronouns and gender identities, it is found that while singular *they* incurs a P600 effect in comparison with plural *they*, it does not show a frontal negativity effect which is present in “gender mismatch” sentences where e.g. *he* refers to referent introduced by an antecedent with strong feminine conceptual gender bias, indicating reference resolution problems. The study also reports a reduction in P600 in older students and suggests that more time spent in the progressive university environment leads to eased processing of singular *they*. Further, comparing ERP measures from students that provided higher offline acceptability rating scores for singular *they* with named antecedents to students with low ratings, no difference between groups is found. The authors conclude that singular *they* with named antecedents incurs difficulty with processing but not reference resolution even in speakers who accept this use, but increased exposure may ameliorate the processing difficulty.

Overall, I believe it has been shown that more recent psycholinguistic scholarship has sometimes failed to employ a consistent trans linguistics approach, despite the groundbreaking work that has shown both the necessity and benefit of such an approach (Ackerman, 2018b; Conrod, 2019; Konnelly & Cowper, 2020). I now turn to production studies of singular *they*. While the majority of production studies have drawn their data from corpora, I will review studies that involve experimental manipulation in the following.
2.2.4. Production Studies

Hekanaho (2020) presents a holistic account of epicene singular *they* and neo-pronouns (neologistic pronouns such as *ze* or *xe*), drawing from rating and production data, as well as usage and attitude surveys. Out of over 1100 participants from North America, Australia, the UK, Finland and Sweden, 9% are trans, of which 79 are non-binary. In an open-ended writing production task, it is found that singular *they* is the most commonly preferred generic pronoun; older people, Finnish people and cisgender men are more likely to use gendered pronouns. Older people and Finns are also more likely to use gendered pronouns in a written cloze test (fill the gap) task of generic pronouns; however, cisgender women are more likely to use gendered pronouns than cis men. Comparing English-speaking countries, British and Australian people are more likely to use nongendered pronouns compared to US-Americans and Canadians. Further, trans* people are significantly more likely to use non-gendered pronouns, as are people who hold negative attitudes towards sexist language. It is also reported that, with some exceptions, usage and acceptability ratings of generic singular *they* largely align within individual participants. Querying acceptability of *they* as a singular pronoun, higher acceptability ratings are found in younger participants, trans* participants, those who self-identify as politically liberal, support non-sexist language, accept transgender people or report knowing transgender people. Within the English-speaking countries sample, only Canadians deviate significantly, rating singular pronoun *they* significantly lower than Britons, Australians and US-Americans.

An innovative written production and memory study of personal pronoun *they* finds that *they* is both remembered and produced less well than binary (*she* or *he*) pronouns (Gardner & Brown-Schmidt, 2021). Participants are exposed to a Public Service Announcement (PSA) about gender-neutral language and/or a short biography including *they/he/she* pronouns, before receiving more information (including pronouns) about a particular referent, followed by an un-related distractor task. Subsequent memory and production elicitation tasks reveal that singular *they* pronouns are remembered less well than *he* or *she*, but this effect is slightly reduced through exposure to the pronoun PSA. The same pattern emerges in production, but production of singular *they* overall is less accurate than memory. Further, exposure to the short biographies implicitly containing *they/them* pronouns did not significantly improve production, but production increased if participants heard both the biographies and the PSA. In another study, Gardner and Brown-Schmidt (2023) find that oral production of singular
they is significantly higher if nametags are provided and/or if people’s pronouns are mentioned as part of an introduction. Moreover, participants’ beliefs regarding the gender binary and gender essentialism presented the strongest predictor for accurate production of personal pronoun they.

One of the relatively few studies that have treated both oral and written production of singular they through experimental manipulation reports increased production if referents’ pronouns are provided (Kramer et al., 2022). The oral production task features a sample of younger speakers while the written production task sample includes more varied ages, and both samples consist of US-American English language users. In the oral production experiment, participants are shown pictures of people with feminine, masculine or androgynous gender presentation and are asked to produce short narratives about the depicted person, using provided verbs. Afterwards, they are asked to do the same but this time, pronouns are also provided alongside the pictures. The written production task was identical, except that pronouns in the second half of the task were only available at the beginning and not for the whole duration of the task. The results show a significant increase of they-production in both oral and written production task if pronouns are available; however, this increase was much larger in the oral task. Across both tasks, people were less likely to correctly gender androgynously-presenting people than those presenting as feminine or masculine. Only in the writing task was there a significant effect of participants’ gender as well as their familiarity with trans*, non-binary and queer people. Men were significantly less likely to use singular they if pronouns were provided, while participants with higher familiarity scores used singular they significantly more in the androgynous condition than less familiar participants. Participants also used singular they significantly more with androgynous referents, regardless of whether pronouns were provided or not. There was no effect of participant age. While Kramer et al. provide an interesting insight into oral and written production of singular they in light of both language-internal and -external factors, they note the relatively small numbers of older as well as non-binary and trans* participants.

The last study of this review presents a trans linguistics approach to the production and processing of singular they, as well as neopronouns, finding no significant disadvantage in reading and writing they compared to she and he (Callaway, 2022). Surveying the pronoun practices of over 1500 (mostly US-American) genderqueer/nonbinary (GQNB) people, it is found that almost three quarters of participants use they/them pronouns exclusively or along
with other pronouns. Further, it also emerges that participants are forced to change which pronouns they go by depending on the space they are in, i.e. more conservative spaces compared to more accepting and progressive ones. Pronoun processing is tested using a stationary window reading task while production is tested by means of a written cloze test afterwards. The reading task contained a short narrative paragraph that featured the pronoun in all its form, e.g. *they*, *them*, *their*, which were then tested in the cloze test. 90 (again mostly US-American) participants are recruited for GQNB familiarity – being GQNB themselves, knowing someone who is GQNB or not personally knowing any GQNB people – and overall there is a roughly even split between cisgender and non-cisgender participants. In both reading and writing, singular *they* does not show a significant disadvantage compared to binary third person pronouns. Different demographic variables such as age, gender, native language as well as familiarity with GQNB people and pronouns do not show a significant effect on reading times. In production, familiarity with GQNB predicts correct responses more reliably than familiarity with different pronouns, which may hint at ideological rather than frequency-based drivers behind pronoun language change. Further, more instances of misgendering are observed in production of *they* and *it* compared to neo-pronouns, suggesting that pronouns that are more popular and thus widely known to index GQNB identities are prone to ideologically motivated misgendering rather than due to memory recall error.

2.3. Motivation for Own Study

I hope this literature review has shown that while a gradual shift towards trans linguistics approaches can be observed in the study of singular *they*, there is still a long way to go. Especially works focussing on processing often lack sociolinguistic approaches, assuming that variation in processing can be attributed to language-internal factors only. Even in studies that do apply trans linguistics frameworks, there is often a relatively small number of trans* and/or non-binary participants. Beyond the area of recruitment, cis- and heteronormative assumptions still pervade the discipline, subconsciously or not. Authors may explicitly or implicitly categorise *they* as a plural pronoun and posit a “number mismatch” if it appears in singular contexts. Another symptom is the positing of a “gender mismatch”, e.g. if *she* refers to an antecedent that has “traditionally” (i.e. from a cisgender perspective) invoked masculine conceptual gender. These assumptions do not just present theoretical linguistic approaches, they mirror biases and prejudices that gender-diverse people encounter
in the real world, which makes it all the more crucial to reflect upon and, in my opinion, eliminate them.

Furthermore, it seems that the majority of studies on singular they, and particularly those that propose underlying syntactic mechanisms, utilise rating data and/or reading time data, while there seems to be relatively less published experimental work on they-production, particularly oral production. And finally, most studies feature a small variety of singular they antecedents, often distinguishing them either syntactically or pragmatically. In contrast, my study aims to provide a holistic picture of rating, reading and production data across a wide range of antecedents. By having all participants provide all three data types, it will be possible to identify any correlations between rating, reading and speaking patterns.

I therefore set out to collect rating, reading and oral production of singular they across syntactically and pragmatically varied antecedents, and apply a trans linguistics approach to answer the following questions:

1. Which language-internal and -external factors influence perception and production of singular they? Is the current language change being driven by language user’s mental concept of gender, or is it simply a question of frequent exposure to singular they?
2. Do rating, reading and production data show the same patterns?
3. Does a trans linguistic approach to the study of singular they lead to new insights?

2.4. Predictions

Based on the above reviewed literature, the following patterns are expected to emerge. In general, some asymmetries between data types are expected, with rating and production data showing significant effects of demographic factors (Hekanaho, 2020) but a lack of such effects in reading time data (Block, 2019). Even though British language users have been reported to produce epicene singular they at a higher rate than US-Americans (Hekanaho, 2020), there have been no studies explicitly comparing differences in significant effects of language-external or -internal factors in the two varieties, so I will rely on previous results to form predictions regardless of participant residence. I will break down predictions by data type in the following.
If British English is undergoing a similar language change to US-American English, naturalness ratings are likely to show an age effect, with older speakers rating singular they less natural than younger speakers (Conrod, 2019). This change may also manifest in the emergence of three different stages/groups of language users according to Bjorkman (2017) and Konnelly and Cowper (2020); in this case, quantified and indefinite epicene antecedents should receive higher naturalness ratings than specific definite antecedents, who in turn should be rated higher than gender-biased NP and given name antecedents. However, antecedent-based effects are likely to be absent within trans* participants (Conrod, 2019), with trans* and non-binary people rating sentences significantly higher than cis people/binary gender people (Block, 2019; Conrod, 2019). Further, if attitudes towards singular they are influenced by real-life experience with people who go by they/them pronouns, participants who are more familiar with non-binary people are expected to provide higher ratings overall (Ackerman, 2018a; Camilliere et al., 2021).

For reading time measures, the main and perhaps only significant effect is predicted to be caused by antecedent condition. Specifically, if Han and Moulton (2022)’s processing account is upheld, gender-biased NP and gender-biased given name antecedents are expected to incur significantly higher reading times than antecedents without gender bias. No significant effects of participants’ (non-binary) gender or frequent exposure to non-binary people are expected to emerge (Block, 2019), although marginal frequency effects may be found (Ackerman, 2018a). However, if the gender-diverse sampling reveals previously unreported demographic effects, this may provide more insight into how language users’ mental concepts shape linguistic processing.

As for the production data, it is somewhat harder to predict outcomes due to sparse previous literature of oral production of singular they after priming. The following predictions are tentative only. There may be a positive priming effect within the gender-neutral name antecedent condition, but no effect of participant gender or frequent interaction with non-binary people (Kramer et al., 2022). However, it may be that oral production data largely patterns with written production, in which case there may be no effect of priming (Gardner & Brown-Schmidt, 2021) but instead age and gender effects, with more singular they-production in younger and gender-diverse language users (Hekanaho, 2020). It is hoped that findings will indicate underlying drivers that may advance the current language change,
thereby helping to reduce instances of misgendering of people who go by *they/them* pronouns.
3. Methods

3.1. Materials

The main focus of the language-internal manipulation was to compare six different antecedent conditions based on previous studies on singular they, which so far have included subsets of the conditions analysed in this study. Bjorkman (2017) identifies two groups of language users based on their acceptance of the pronoun: conservative users, who accept bound singular they with quantified (I) and indefinite epicene NPs (II), and innovative users, who also accept it with definite, specific NPs that lack strong binary conceptual gender connotations (III). Konnelly and Cowper (2020) confirm the existence of a third group, or stage, with language users accepting singular they with any antecedent, including definite NPs with “traditionally” binary conceptual gender association (IV) as well as given names. However, Ackerman (2018a) observes variation in acceptability and processing depending on the gender association evoked by names, hence I separate names with no strong gender association (V) from those with traditionally masculine or feminine associated names (VI). Gender-neutral and gender-biased nouns were chosen from previous norming studies (Doherty & Conklin, 2017; Kennison & Trofe, 2003), whereas gender associations of given names were tested in a norming study that will be described further on.

(I) Every doctor tries their best every day.
(II) A great doctor tries their best every day.
(III) This doctor tries their best every day.
(IV) The girl tries their best every day.
(V) Alex tries their best every day.
(VI) Robert tries their best every day.

All test items in the naturalness rating and in the reading time task featured the possessive their form to minimise variation. Their was chosen over other forms due to the following reasons. Nominative they would require a previous main clause introducing the antecedent, which would have increased the already considerable length of the experiment. Accusative object them would require an additional subject denoting a different referent in the same sentence, which could potentially lead to reference ambiguity and favouring of a plural reading of they. As for the reflexive, there are currently two forms possible for singular they –
themselves or themself – and studies have found significant differences in ratings of the two different forms (Ackerman et al., 2018; Conrod et al., 2022).

In order to increase the likelihood of a singular reading of they the following techniques were employed: the use of formulaic phrases such as “try one’s best” or “fear for one’s safety”, use of own and self to “glue” the pronoun to the subject, e.g. in “listen to one’s own voice” or “self-publish one’s book”, use of phrases containing body parts such as in “crane one’s neck” or “donate one’s kidney” to signify physical attachment, as well as use of adjuncts containing an action that reinforced singular reading such as “stretch one’s legs before running” or “meet one’s hero before dying”. The high percentage of correct answers (>80%) to the comprehension questions, which focused on singular interpretation, suggests that these strategies were mostly successful. A full list of materials can be viewed in the Appendix.

3.2. General Set-Up

All experiments were approved by the Research Ethics Committee of the School of Philosophy, Psychology and Language Sciences (PPLS) of the University of Edinburgh on 10th March 2022, with minor amendments accepted on 22nd July 2022 (reference number 217-2122/4).

Participants first read an information sheet, briefing them about the nature of the study, compensation, data use and protection and the right to withdraw at any point. After reading the information sheet, participants then read a consent sheet and confirmed their consent by clicking on a button leading them to the experiment. Information sheet and consent form templates were used as provided by the Linguistics and English Language research ethics information webpage (Linguistics and English Language Research Ethics Information - Consent form templates, 2018).

Further, participants were asked to complete the experiment in a quiet environment and encouraged to use headphones and microphones. Throughout the experiment, a progress bar was shown at the top bar to provide an accessible overview. Participants were encouraged to complete the experiment swiftly and to take breaks in-between tasks if needed.
3.3. General Data Management

Raw data containing personal identifiers were encrypted and stored on the DataStore repository. This data was then pseudonymised, and the pseudonymised versions were stored on OneDrive. The participant keys linking pseudonymised to raw data were also encrypted and stored on DataStore, separate from the raw data. Audio recordings were also encrypted and stored on DataStore. Audio files were coded for pronoun use, and the coded data was stored on OneDrive. Upon completion of this dissertation, the raw data will be transferred to the supervisors.

3.4. Norming Study

A norming study was carried out to determine gender biases of given names. A total of 40 names were rated on a Likert-scale with labelled end points: 1 (masculine) to 7 (feminine). Each participant provided a rating for each name, and at the end provided optional demographic information which is detailed below in Table 1.

Out of the 40 name items, 10 were selected to have masculine bias, 10 with feminine bias, and 20 to be gender-neutral. Masculine and feminine names were picked from the National Records of Scotland’s data on popular baby names, spanning 1950 to 2020, and gender-neutral names were selected from data from 1974 – 2020 (Dixon, 2021; Jackson & Donnelly, 2001; The Most Popular Names in Scotland, 1998, 1999). The smaller time frame for gender-neutral names is due to 1974 marking the year from when the data included the quantity of babies named, whereas records prior to then only listed the names and whether they were given to babies assigned male or female at birth. Gender-neutral names, on the other hand, were selected by picking popular names that were given to a roughly equal number of babies assigned female and those assigned male at birth. Due to the predicted greater variety, more gender-neutral names were included in the norming study than masculine and feminine names.

The norming study was hosted on the experiment platform “Testable” (Rezlescu et al., 2020), and participants were recruited through Prolific (www.prolific.co) [14/04/2022]. As the experiment was predicted to take at most 10 minutes, participants were compensated £1.59 in line with Scottish Living Wage at the time (£9.54/hour). The study was launched and completed by 30 participants who were located in the UK. Participant ages ranged from 18 to
66 years, with a mean of 30 and a standard deviation (SD) of 12.45. A detailed breakdown of participant demographics is shown in Table 1. The initial aim was to include an even number of trans* and cisgender participants, but due to allowing people who did not disclose to Prolific whether they are trans* or cis, this resulted in a higher number of cis participants. Further, participants were chosen if they had acquired English before the age of five. Results of the norming study are included in the appendix.¹

**Table 1**

*Demographic Breakdown of Norming Study Participants*

<table>
<thead>
<tr>
<th>Demographic Variable</th>
<th>Levels (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sexuality</td>
<td>Heterosexual (18), bisexual (4), asexual (3), pansexual (2), lesbian (2), queer (1)</td>
</tr>
<tr>
<td>Gender</td>
<td>Woman (15), man (8), non-binary (4), genderqueer (1), genderfluid (1), agender (1)</td>
</tr>
<tr>
<td>Pronouns</td>
<td><em>She/her</em> (15), <em>he/him</em> (8), <em>they/she</em> or <em>she/they</em> (4), <em>they/he</em> (1), <em>they/them</em> (1), any (1)</td>
</tr>
<tr>
<td>Trans* identity</td>
<td>Yes (5), No (24), Prefer not to say (1)</td>
</tr>
<tr>
<td>Taken a linguistics course at degree level</td>
<td>Yes (3), No (27)</td>
</tr>
</tbody>
</table>

**3.5. Pilot Study**

A pilot study was set up on Testable (Rezlescu et al., 2020) in order to gauge average completion time and gather participant feedback for improvement. Participants were recruited through Prolific (www.prolific.co) [18/05/2022] (n = 4) as well as through UserTesting (www.usertesting.com) (n = 11). UserTesting automatically records the screen and audio of participants as they review websites, but this data was not stored for the purposes of the pilot study. So-called “testers” are paid USD$10 per test, whereas Prolific participants were paid £6.34 in line with Scottish Living Wage at the time.

¹ Preliminary analysis indicated that younger people, trans* people and those who go by *they/them* or multiple pronouns are less likely to provide ratings at extreme ends of the scale. However, this was not further explored due to the small sample size and it lying outwith the remit of this study.
The main feedback was the longer than expected length of the experiment, so a disclaimer page was added. In the naturalness rating task, some participants remarked they wished to go back and amend previous ratings after encountering ungrammatical filler items, so one Standard English and one ungrammatical filler item were moved to the beginning of the task to expose participants to the extreme ends of the scale. The data from the production task yielded a lot of unusable data as participants were struggling to understand what they were required to do, so the number of examples in the training phase was increased from one to three. Finally, and perhaps most importantly, one participant expressed anxiety regarding the aims of the study, noting the vulnerability of gender-variant people and distrust towards scientific research due to historical injustices. In order to reassure participants as well as to educate those less familiar with singular they, resources about pronouns and LGBTQ+ allyship were added to the debrief.

3.6. Main study

3.6.1. Participant Recruitment

Participants were recruited through several means. First, calls were posted to local Facebook groups (Queer Edinburgh, Meadows Share, LGBT+ Students at the University of Edinburgh, GLASGOW LGBTQIA+ COMMUNITY, Trans Active Edinburgh, Yakety Yak Language Cafe) as well as the University of Edinburgh’s Staff Pride Network mailing list, in order to receive local responses.

Prolific (www.prolific.co) [25/01/2023] was then used to screen for demographic groups that were underrepresented in the local responses. However, some issues emerged with the internal pre-screening tool as some Prolific users had reported their gender identity to differ from their sex assigned at birth but did not reflect this in the demographic questionnaire. When this was followed up, most of these users admitted they were cisgender and misread the Prolific screening question, so their submissions were rejected. I also used the University of Edinburgh’s PPLS Sona system to recruit student participants; as this system does not allow pre-screening selection, this yielded a lot of “unwanted” participants who are not included in the main analysis. Further, due to some technical issues with the Testable website, audio data for some participants was distorted beyond recognition nor recovery, and to make up for this, Testable provided credit to recruit additional participants through their Testable Minds platform.
3.6.2. Participant Compensation

Participants recruited through social media and mailing lists were sent a “One4all” gift voucher while the remaining participants were paid via their recruitment platform. While the average completion time for the pilot study was 37 minutes, it was not possible to send gift vouchers with amounts in pence, so the payment was rounded up to £7 per participant. Student participants recruited through Sona systems were rewarded 1 course credit for their participation.

3.6.3. Structure

The final study consisted of three separate experiments and a final demographics questionnaire. I will describe the individual components in the following; antecedent conditions are repeated below for ease of reference.

(I) Quantified gender-neutral NP
(II) Indefinite epicene gender-neutral NP
(III) Definite specific gender-neutral NP
(IV) Definite gender-biased NP
(V) Gender-neutral name
(VI) Gender-biased name

1: Rating Task

The naturalness rating task was set up to compare ratings of antecedent conditions that had been previously studied in separate works (e.g. equivalents of AC I-III in Bjorkman (2017), AC IV-VI in Konnelly and Cowper (2020), and AC I, III, V and VI in Conrod (2019)) . This task required participants to rate one sentence at a time, on a 6-point Likert scale with labelled end points (1 – unnatural, 6 – natural). The labels, as well as instructions, focus on naturalness to convey the subjectivity of ratings, rather than an imagined authority that may be evoked with grammaticality or acceptability ratings. This is especially important to consider with regards to linguistic judgements being extrapolated to judgements of the people who use the language being judged, and in the case of singular they care must be taken to avoid judgement of trans* and non-binary people who use the pronoun (Ackerman, 2018b).
A six-point scale was chosen in order to avoid a middle neutral point, and to force participants to decide whether they found a sentence more natural or more unnatural.

Participants were asked to provide two ratings per sentence, one when imagining if the sentence was part of their own speech, and the other one imagining they were reacting to someone else’s speech. This was done to control for cases where people may not use singular they but may not have an issue with other people using it. Further, participants were instructed to rate how natural the sentence sounds rather than how much they agree or disagree with its content, and not to overthink their answer as there were no right or wrong outcomes.

Test sentences contained all six antecedent conditions (ACs) described in the “3.1. Materials” section. However, each participant saw only one item per AC – for example, if they saw the doctor item in the quantified AC, they would see a different item in the indefinite epicene AC and so on. Overall, there were six AC and thus six items to allow for even distribution, so 36 sentences in total. Participants were randomly allocated to one of six groups, each of which contained six sentences. Further, participants also rated the same 12 filler sentences, as per the recommended 1:2 ratio of test to filler material (Cowart, 1997). To provide a variety of naturalness across fillers, six of them were Standard English, four contained non-standard lexical items or morphosyntax, and two were ungrammatical. The task started with one standard filler, followed by an ungrammatical filler, and the remaining sentences were presented in randomised order.

2: Reading Task

The aim of the reading experiment was to collect reading time data that could be compared to the naturalness rating data of the first task. Most previous studies on singular they have focused on either rating or reading data, and the studies that combine both data types have tended to survey a more limited number of different antecedent types (e.g. indefinite epicene NP [AC II] of varying gender expectancy in Doherty and Conklin (2017) and Block (2019), names of varying gender association [AC V-VI] and indefinite pronouns in Ackerman (2018a)).

In this second task, one sentence at a time was presented on the screen. Participants were asked to read silently at their usual speed, and to press the space bar once they had
comprehended the meaning of the sentence. A quarter of sentences were followed by comprehension questions to test for singular interpretation of *their*, as well as to avoid participants from skimming or skipping sentences.

The materials followed the same set-up as in the rating task, and participants remained in the same six groups. However, in addition to the 6x6 set that was rated in the first part, three additional sets were added to control for idiosyncratic variation across items. In total, participants read 24 test sentences – four sentences for each antecedent condition. The number of fillers was also quadrupled to 48, with 24 Standard English, 16 non-standard and 8 ungrammatical fillers.

3: Speaking Task

The oral production task presented a novel exploration of the impact of auditory priming on singular *they* production, as previous experimental production studies have examined written production (Callaway, 2022; Gardner & Brown-Schmidt, 2021; Hekanaho, 2020) or also provided visual depictions of referents (Kramer et al., 2022). The production task in this study was split in half into a listening and speaking phase, and closely modelled after Kaschak and Glenberg (2004)’s syntactic priming experiment.

After completing the reading task, participants were asked to take a short break of 1 minute before proceeding to the third task. In the listening phase, participants were instructed to listen to ten short stories and to answer one question about the content after each story. Participants were randomly assigned to either a test or control group. They listened to five short test or control stories, each five sentences long, containing short narratives about two people. The two referents both corresponded to antecedent conditions IV – VI, so there was one story each containing two referents that were denoted by either masculine-biased NPs, feminine-biased NPs, masculine-biased given names, feminine-biased names or gender-neutral names. AC IV-VI were chosen due to their relatively variable status across the three stages of singular *they* language change, as it was assumed that AC I-III were more likely to already produce a ceiling effect and be less susceptible to possible priming.

In the test group, one of the referents was referred to with *they/them* pronouns, whereas in the control group, one of the referents was referred to with either *she/her* or *he/him* pronouns. The second referent was not referred to with any third-person pronouns. Additional to these
five test or control stories, both groups listened to the same five filler items that did not contain singular *they, she* or *he* pronouns.

After each story, participants had to answer a Yes/No comprehension question that was designed to act as a distractor from the pronoun. Correct answers were evenly split with five “yes” and five “no” answers. After the listening phase was complete, participants were asked to wait for two minutes before moving on to the speaking phase, with a 2-minute countdown displayed on screen for reference.

In the final part of the experiments, participants saw one direct speech sentence at a time, and were asked to transform it into indirect speech. By doing so, they would have to transform a first-person singular pronoun (9) into third person (10), thus choosing between *he/she/they*. The instructions concluded a plural example so as not to influence the critical singular items.

(9) Emily says: “I need to stop grinding my teeth.”

(10) Emily says that [he/she/they] need[-s] to stop grinding [his/her/their] teeth.

A 30-second timer indicated how much time participants should spend per sentence, although they were able to spend more time if needed. Once an audio signal was detected, audio was automatically recorded for ten seconds before automatically progressing to the next sentence. The task started with three practice sentences, after which participants saw the solution. These practice sentences contained either plural, second person, or inanimate subjects, again to avoid influencing critical third-person singular items following the practice runs.

The speakers and thus referents in these sentences were identical to the ones used in the short stories of the listening task, totalling ten critical sentences. In addition, there were 20 filler sentences, nine of which contained distractors in the form of non-anglophone given names (Campbell, 1996).

4: Demographics and Attitudes

After the four main experiments were completed, participants were asked rate a statement about gender (11), a statement about “biological sex” (12), and to rate a sentence containing
singular *they* with a masculine-biased given name, as well as providing reasons for each of these three ratings. This data was collected in order to yield a very broad idea of participant’s explicit and implicit attitude towards gender-diverse people and pronouns.

(11) The gender binary is a myth: gender exists on a spectrum.

(12) You can't tell what sex a person is by looking at them.

After this, participants were asked to provide optional demographic data of interest for the analysis. Sexuality, gender and pronouns were collected through the use of open text boxes, taking guidance from Conrod (2021). Other demographic information collected included participants’ age, how often participants interact with people that go by *they/them* pronouns, whether participants identify as part of the trans* umbrella, whether they acquired the English language before or after the age of 5 and whether they had ever taken a Linguistics course at degree level. All questions were optional and those with pre-defined answers included a “prefer not to say” level.

Finally, participants could leave general comments in an open text box. After this was completed or skipped, participants were invited to visit links leading to the APA Style guide on singular *they*, pronouns.org as well as to Stonewall’s and Amnesty International’s web resources on gender identity and trans* allyship.
4. Analysis and Results

4.1. Participant Demographic Overview

Although demographic data was collected at the end of the study, I will begin this section with an overview of the participant sample. In total, 120 people took part in the study. After excluding those who reported learning English after the age of 5 or having taken a course in linguistics at degree level, as well as any who selected “Prefer not to say” for any of the relevant demographic measure, 65 participants remained.

Despite the use of various recruitment platforms, it proved rather hard to recruit a 1:1 ratio of trans* and cisgender participant. In the end, I opted for a roughly 50% share of participants who do at least one of the following: (a) consider themselves part of the trans* umbrella, (b) report their gender as neither male/man nor female/woman or (c) go by they/them or multiple different pronouns. The ages at time of study completion range from 18 to 67 years, with a mean of 28.85 and a standard deviation (henceforth SD) of 12.21. A detailed breakdown of further demographic and other participant measures can be seen in Table 2. All analysis was conducted using R in RStudio (R Core Team, 2022).

Table 2
Demographic Overview of Main Study Participants

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>R coding</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age in years</td>
<td>age</td>
<td>18-67, mean 28.85, standard deviation (SD) 12.21</td>
</tr>
<tr>
<td>Gender</td>
<td>gender</td>
<td>30 women, 15 men, 13 non-binary, 1 genderfluid, 1 genderqueer, 1 genderqueer man, 1 genderqueer (/nonbinary), 1 agender, 1 non-binary (agender)</td>
</tr>
<tr>
<td>Part of trans* umbrella</td>
<td>trans</td>
<td>39 no, 26 yes</td>
</tr>
<tr>
<td>Participant’s pronoun(s)</td>
<td>pronouns</td>
<td>29 she/her, 15 he/him, 14 they/them, 3 she/they, 1 he/it, 1 they/he, 1 they/she, 1 she/he/they</td>
</tr>
</tbody>
</table>
Sexuality

30 heterosexual, 13 bisexual, 9 queer, 2 asexual, 2 pansexual, 2 lesbian, 1 gay, 1 omnisexual, 1 bisexual/pansexual, 1 asexual/heterosexual, 1 aromantic and queer, 1 aromantic pansexual, 1 gay/queer

Interaction with people who go by they/them

14 daily, 25 weekly, 10 monthly, 5 yearly, 11 never

Due to the high number of factorial independent variables, I first undertook Factorial Analysis of Mixed Data (FAMD), a generalisation of Principal Component Analysis as developed by Blaufuks (2021). However, FAMD showed that the first two principal components together only accounted for 39.6% of variance. I therefore calculated the Variance Inflation Factor (VIF), included in the car package (Fox & Weisberg, 2019). In order to calculate the VIF, a maximum Linear Mixed Effects model was fitted:

\[ \text{lm} \text{er}(\text{rating} \sim \text{condition} + \text{pronouns} + \text{they}_\text{frequency} + \text{rate}_\text{sex}_\text{statement} + \text{age} + \text{gender} + \text{trans} + \text{sexuality} + \text{speech} + (1 \mid \text{Participant.ID})) \]

The resulting variance-inflation factors indicated high collinearity for gender (4784), sexuality (834) and pronouns (471), with the next-highest factor being they_frequency (12). In order to circumvent collinearity issues, I decided to include only pronouns in the model; they can be seen as a stand-in for gender and sexuality in this sample (see Table 3).

Further, due to an overall small sample, some levels within the independent variables had very few members, which led me to collapse them. In the pronouns variable, I decided to reduce the levels to four: he/him, she/her, they/them, and “multiple”, the latter of which combined participants that use more than one pronoun form e.g. she and they pronouns. This is potentially problematic as people may be using multiple pronouns for different reasons; one person may prefer to be referred to using differing pronouns regardless of context, while others may ideally go by one particular pronoun but also have more normative pronouns for use in contexts where they may not feel safe or comfortable enough to go by their first-choice pronoun. Future research should seek to distinguish these and possibly further differing pronoun patterns. Further, the last two levels within they_frequency were collapsed into an
“almost never” level due to low numbers as well as the relatively small difference between interacting with one compared to zero they/them-users a year.

Table 3 shows the number of participants by pronouns and gender, with number of trans* participants included in brackets (n*). We can see that all men, and only men, used he/him pronouns, and 28 out of 29 she/her-users are women. Out of the non-binary participants, only one uses binary pronouns exclusively (she/her) with the remainder using they/them as their only or one of their pronouns. All participants who are neither men, women nor non-binary also use they/them as at least one of their pronouns. In contrast, trans* participants are distributed across all gender and pronoun groups. Considering sexuality, the vast majority of heterosexuals (n = 30) were either men (12) or women (17) and used either he/him (12) or she/her (15) pronouns. Further, most heterosexuals were cisgender (26).

Table 3
Participant Breakdown by Personal Pronouns, Gender, and Trans* Identity

<table>
<thead>
<tr>
<th>pronouns/ gender</th>
<th>he/him</th>
<th>she/her</th>
<th>they/them</th>
<th>multiple</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>man</td>
<td>15 (6*)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15 (6*)</td>
</tr>
<tr>
<td>woman</td>
<td>0</td>
<td>28 (5*)</td>
<td>0</td>
<td>2</td>
<td>30 (5*)</td>
</tr>
<tr>
<td>non-binary</td>
<td>0</td>
<td>1</td>
<td>9 (8*)</td>
<td>3 (1*)</td>
<td>13 (9*)</td>
</tr>
<tr>
<td>other</td>
<td>0</td>
<td>0</td>
<td>5 (4*)</td>
<td>2 (2*)</td>
<td>7 (6*)</td>
</tr>
<tr>
<td>total</td>
<td>15 (6*)</td>
<td>29 (5*)</td>
<td>14 (12*)</td>
<td>7 (3*)</td>
<td>65 (26*)</td>
</tr>
</tbody>
</table>

In regard to the progressive statement on gender, participants mostly agreed with a mean rating of 5.86 (SD 1.57) in a Likert scale from 1 (disagree) to 7 (agree). The statement gauging participants’ views on “biological sex” was rated less progressively (mean 4.94, SD 1.94). Reviewing the accompanying comments stating the reason for providing a particular rating, it was revealed that several participants accidentally inverted the scale, most probably due to the statement containing negation. Due to these scale inversions as well as the very progressive skew of gender attitudes, it was decided not to include these measures in the statistical analysis.
4.2. Naturalness Rating Data

Average naturalness ratings of filler sentences were used to check for scale inversions. Two participants were excluded due to rating standard filler sentences on average below the scale mid-point (3.5), and one participant was excluded due to an average rating of ungrammatical filler sentences above the mid-point. After exclusions were applied, the remaining 62 participants had an average rating of 5.56 for standard fillers (SD 0.87). Sentences containing singular *they* were rated somewhat lower, with an average of 4.75 (SD 1.5), but well above non-standard fillers (mean 2.76, SD 1.81) and ungrammatical fillers (mean 1.16, SD 0.51).

An initial look at the data does not contradict the predicted pattern of highest ratings of quantificational and indefinite epicene AC, followed by definite specific NP, with gender-biased NP and both gender-biased and gender-neutral given names with lowest overall ratings (Figure 1). Interestingly, gender-biased NP are rated the lowest by far. However, all conditions are rated above the mid-point of the scale, confirming an overall progressive sample. There were five exceptions to the predicted implicational hierarchy: two participants who rated quantificational and indefinite epicene AC below 3.5, and three participants rating definite specific NP below 3.5 despite rating the “super-innovative” conditions above 3.5. However, these instances do not necessarily constitute a deviation from the three hypothesised stages of singular *they* innovation, as participants only rated one sentence per condition. Further, while I take the mid-point of the scale as the dividing point between natural and unnatural, this interpretation may differ from person to person.
I used the Cumulative Link Mixed Model (\texttt{clmm}) from the \texttt{ordinal} package (Christensen, 2022). While a Bayesian approach was attempted, this had to be abandoned due to limitations in time and resources needed to acquire the required skills. However, \texttt{clmm} preserves the ordinal nature of the outcome variable. Further, I ensured to specify all predictor variables except for age to be factors, with \texttt{they\_frequency} having ordered levels (\texttt{daily} > \texttt{weekly} > \texttt{monthly} > \texttt{almost never}). I fitted a maximum model with random effects by participant and item, main effects of AC, participant’s pronouns, interaction with \texttt{they/them}-users, age, trans*ness and their interaction with AC, as well as a main effect of rating one’s own versus someone else’s speech:
I then removed one effect or interaction at a time, and compared the maximum model with each reduced model using the `anova` function from the `stats` package (R Core Team, 2022) to identify which effects and interactions significantly improved the model. In order to remove main effects of a factor that featured in an interaction, it was necessary to sum-code these factors. Unfortunately, this meant that it was not possible to locate significant effects in the individual level(s), so I will follow up significant effects with a visual exploration. Table 4 below shows the outcomes for `anova` comparisons, with the first column specifying the main effect/interaction that was excluded from the smaller model.

Table 4

Rating Model Comparisons Output

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Likelihood Ratio Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>condition</td>
<td>1.5504</td>
<td>0.9072</td>
</tr>
<tr>
<td>pronouns</td>
<td>-2.2243</td>
<td>1</td>
</tr>
<tr>
<td>pronouns:condition</td>
<td>38.322</td>
<td>&lt; 0.001 ***</td>
</tr>
<tr>
<td>they_frequency</td>
<td>38.891</td>
<td>&lt; 0.001 ***</td>
</tr>
<tr>
<td>they_frequency:condition</td>
<td>33.271</td>
<td>0.004305 **</td>
</tr>
<tr>
<td>age</td>
<td>-3.9945</td>
<td>1</td>
</tr>
<tr>
<td>age:condition</td>
<td>-1.0166</td>
<td>1</td>
</tr>
<tr>
<td>trans</td>
<td>23.153</td>
<td>&lt; 0.001 ***</td>
</tr>
<tr>
<td>trans:condition</td>
<td>-2.0448</td>
<td>1</td>
</tr>
<tr>
<td>speech</td>
<td>25.243</td>
<td>&lt; 0.001 ***</td>
</tr>
</tbody>
</table>

Rather surprisingly, no significant main effect of antecedent condition or age, nor of their interaction, was found, unlike in previous rating studies (Camilliere et al., 2021; Conrod, 2019; Hekanaho, 2020). The lack of age effect may be attributed to the relative youth of the sample.

While there was no significant effect of participant’s pronouns on overall ratings, the interaction of pronouns with antecedent condition did reach significance. Visualisation of
average ratings by pronouns across AC in Figure 2 shows relatively little variation between pronoun groups in the first three, more “established” AC. More pronounced differences emerge in the stage 3/super-innovative conditions: gender-biased NP, gender-neutral name and gender-biased name AC. Both in the gender-biased NP and the gender-neutral name condition, participants who go by *they/them* and those who go by multiple pronouns provide higher ratings than *she/her-* and *he/him-*users. However, only the *they/them*-group rate singular *they* with gender-biased names remarkably higher than people using binary pronouns, while the multiple-pronoun group provides similar average ratings to the binary pronoun groups. The lowest average rating can be observed for the *he/him*-group rating singular *they* with gender-biased NP antecedents, which resulted in an average rating below the scale’s 3.5 mid-point. Keeping in mind the sample’s collinearity of pronouns with gender, these findings mostly align with Bradley (2020), who finds a similar lack of AC effect with non-binary participants, whereas binary participants rate singular *they* with common noun AC significantly higher than with given names.
Turning to the trans variable, on the other hand, I observe a significant main effect between people who are part of the trans* umbrella and those who are not, but no significant interaction with antecedent condition. Trans* participants rate all sentences containing singular they significantly higher on average (5.10, SD 1.34) than cis participants do (4.52, SD 1.57). This mirrors findings by Hekanaho (2020), who reports generally higher ratings by trans* than cis participants across antecedent conditions. Considering both pronouns and trans*ness, my findings contrast with Conrod (2019), who reports a significant main effect of non-binary gender and a significant interaction of trans* identity with antecedent condition –
namely, a lack of antecedent condition effect in trans* participants while cis participants show significant antecedent effects.

Frequency of interaction with people who use *they/them* pronouns for themselves returned both a significant main effect as well as a significant interaction with AC. Examining the main effect first, Table 5 below shows high average ratings by participants who report weekly and daily interaction, and comparatively lower ratings in the monthly and almost never groups.

Table 5

Mean Naturalness Ratings by Participant’s Self-reported Interaction with People who go by *they/them*

<table>
<thead>
<tr>
<th>Frequency of interaction with <em>they/them</em>-users</th>
<th>Mean naturalness rating</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekly</td>
<td>5.20</td>
<td>1.12</td>
</tr>
<tr>
<td>Daily</td>
<td>4.94</td>
<td>1.37</td>
</tr>
<tr>
<td>Monthly</td>
<td>4.27</td>
<td>1.75</td>
</tr>
<tr>
<td>Almost never</td>
<td>4.16</td>
<td>1.71</td>
</tr>
</tbody>
</table>

Figure 3 further details the interaction between *they_frequency* and antecedent conditions. Similarly to the interaction between pronouns and AC, sizable difference emerge only in the stage 3 conditions (definite gendered NP and both name AC), where the daily and weekly group provides much higher ratings than the other two groups. Interestingly, in the gender-biased NP condition, monthly participants rate singular *they* even lower than the almost never group, but both groups have an average rating below the 3.5 mid-point. In both given name conditions, only the almost never group provides <3.5 ratings. Another noteworthy observation is located in the indefinite epicene NP condition, where daily participants (mean 4.82) rate singular *they* considerably lower than the weekly group (5.60). For sentences with quantified singular *they*, only the monthly group rates them less natural (4.83) than the other three groups (5.40 – 5.50).
Finally, the modelling returned a significant effect of whose speech was being rated. However, the size of this effect is small. Naturalness ratings for one’s own speech averaged 4.68 (SD 1.57) compared to 4.83 (SD 1.44) for someone else’s speech.

4.2.1. Excluded Rating Data

After analysis of the main data, a follow up analysis was conducted to investigate the impact of age of English acquisition (eng) and as well as degree-level linguistics education.
(ling). The maximum model from the main analysis was modified by adding the main effects of these two variables, as well as an interaction with antecedent condition. Again, model comparison through use of anova was employed and as the results in Table 6 show, this comparison returned a significant effect of the interaction between age at English acquisition and AC.

**Table 6**

*Rating Model Comparisons Output with Excluded Participants*

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Likelihood Ratio Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>eng</td>
<td>-24.311</td>
<td>1</td>
</tr>
<tr>
<td>eng:condition</td>
<td>13.419</td>
<td>0.01975 *</td>
</tr>
<tr>
<td>ling</td>
<td>-25.565</td>
<td>1</td>
</tr>
<tr>
<td>ling:condition</td>
<td>-9.8497</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 4 illustrates the differences between antecedent conditions. In the quantified condition as well as in the gender-biased NP condition, differences between the two groups are relatively small. In all other conditions, participants who learned English after the age of 5 rate singular they much less natural than those who learned English before the age of 5.
To analyse the reading time (RT) data, I first divided the total reading time per sentence by the number of characters of the sentence to obtain a comparable measure of millisecond per character (ms/c henceforth). For a general overview of the data, let us consider mean reading times first. Sentences containing singular *they* were read at a speed of 54.34 ms/c on average (SD 60.98), slightly slower compared to standard filler sentences at 47.34 ms/c. The average
RT for non-standard fillers was much slower at 69.94 ms/c (SD 214.1). Ungrammatical filler sentences averaged a reading speed of 113.10 ms/c (SD 80.83).

I then coded the responses to the comprehension questions as correct (1) or incorrect (0) in order to check participants’ attention and also their singular reading of they. No participants had to be excluded, as everyone answered at least 80% of questions correctly, suggesting a strong singular interpretation of singular they in addition to close attention. However, individual RT measures were excluded if they were extremely fast or extremely slow and therefore potential outliers. Assuming a minimum possible reading speed of 50 ms per word, the longest word of the task featured 15 characters, restricting the fastest allowed speed at 3.3 ms/c. As for the slowest allowed speed, I first considered excluding data 10 SD above the mean as in e.g. Han and Moulton (2022). Upon visualising all data points, it became clear that a measure of 3 SD above the mean provided an improved exclusion criteria, thus RT measures above 238 ms/c were excluded (18 out of 1560 data points).

For inferential statistical analysis, I used Linear-Mixed Effects Models (lmer) from the lme4 package (Bates et al., 2015). Again, I fitted a maximum model after sum-coding the categorical variables, and then compared the maximum model to a reduced model lacking one main effect or interaction at a time. The main model included antecedent condition, participants’ pronouns, interaction with they/them-users, age, trans* identity and mean ratings of the previous task, as well as an interaction with condition of all predictors and random effects by participant and by item (15). Table 7 shows anova comparison outcomes below.

\[
(15) \text{lmer}(RT \sim \text{condition}^* \text{pronouns} + \text{they}_\text{frequency} + \text{they}_\text{frequency}:\text{condition} + \text{age} + \text{age}:\text{condition} + \text{trans} + \text{trans}:\text{condition} + \text{rating}_\text{mean} + \text{rating}_\text{mean}:\text{condition} + (1|\text{Participant.ID}) + (1|\text{item}))
\]

Table 7
Reading Time Model Comparisons Output

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Chi-squared</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>condition</td>
<td>4.9368</td>
<td>0.4236</td>
</tr>
<tr>
<td>pronouns</td>
<td>5.181</td>
<td>0.159</td>
</tr>
<tr>
<td>pronouns:condition</td>
<td>26.555</td>
<td>0.03258 *</td>
</tr>
<tr>
<td>they_frequency</td>
<td>5.2746</td>
<td>0.1528</td>
</tr>
</tbody>
</table>
Unlike most previous studies (Ackerman, 2018a; Block, 2019; Foertsch & Gernsbacher, 1997; Han & Moulton, 2022), the analysis did not return a significant effect of antecedent condition. However, its interaction with pronouns did reach significance. Figure 5 shows the average reading time by pronoun group across the six antecedent conditions. In the quantificational, definite specific and gender-neutral name condition, no striking differences emerge between pronoun users. In the indefinite epicene AC, the groups seem somewhat further dispersed, with he/him-users the slowest at an average RT of 52.04 ms/c (SD 28.23), followed by they/them-users with 46.21 ms/c (SD 24.65), she/her-users (mean 43.14 ms/c, SD 22.53) and multiple-pronoun users with the fastest average RT at 39.13 ms/c (SD 15.90). The he/him-group also reads sentences where singular they occurs with a gender-biased given name considerably slower than other pronoun users, with a mean RT of 61.40 ms/c (SD 27.16). The other three pronoun groups read gender-biased name condition sentences on average at least 10ms/c faster, between 45-51 ms/c. A somewhat different picture emerges in the gender-biased NP condition, where not only he/him-users but also they/them-users show much slower speeds (64-66 ms/c) than the she/her- and the multiple-pronoun group (45 – 47 ms/c).
This interaction of participant’s pronouns with antecedent condition shines a new light on existing findings, which so far did not report any significant effect of demographic predictors on singular *they* reading times. This predicted absence of significance is borne out for the predictor variables of age, trans* identity and frequency of interaction with people going by *they/Them* pronouns. Further, mean naturalness ratings provided in the first task do not significantly impact reading speeds either.
4.3.1. Excluded Reading Data

As before, this section considers data from participants who acquired the English language after age 5 as well as those who have taken a Linguistics course at degree level. The same methods were applied, and anova comparison returned a significant main effect of English acquisition age, as detailed in Table 8. Participants who learned English after their fifth birthday had a mean reading time of 70.91 ms/c (SD 37.63), whereas the before-5 group averaged a reading speed of 51.59 ms/c (SD 27.80).

Table 8
Rating Time Model Comparisons Output with Excluded Participants

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Chi-squared</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>eng</td>
<td>15.823</td>
<td>&lt; 0.001 ***</td>
</tr>
<tr>
<td>eng:condition</td>
<td>2.8119</td>
<td>0.729</td>
</tr>
<tr>
<td>ling</td>
<td>2.7228</td>
<td>0.09892</td>
</tr>
<tr>
<td>ling:condition</td>
<td>9.4229</td>
<td>0.09334</td>
</tr>
</tbody>
</table>

4.4. Oral Production Data

For the oral production experiment, responses were coded as “1” if participants produced a form of singular they, “0” if they produced forms of he or she, and NA if they misunderstood the task or if their audio data was corrupted. Out of 650 total data points, 95 NAs had to be excluded.

(16) glmer(pronounUsed ~ condition * prime + trans* trans:prime + trans:condition + pronouns + pronouns:prime + they_frequency + age + age:prime + age:condition + (1|Participant.ID),
family = binomial,
control=glmerControl(optimizer="bobyqa",optCtrl=list(maxfun=2e5)))

After sum-coding the factorial independent variables, a maximum model was fitted using a Generalised Linear Mixed-Effects Model (glmer, “ordinal” package (Bates et al., 2015)). The family was set to binomial due to the binary outcome variable, and the “bobyqa” optimiser was used to counter convergence issues. The maximum model contained main effects of antecedent condition, priming group, trans* identity, pronouns, age, frequency of
interaction with they/them-users as well as various interactions and a random effect by participant, as detailed in (16). Again, this was compared to reduced models missing one effect or interaction at a time. The results are summarised in Table 9.

Table 9
Production Model Comparisons Output

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Chi-squared</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>condition</td>
<td>24.49</td>
<td>&lt; 0.001 ***</td>
</tr>
<tr>
<td>prime</td>
<td>1.7483</td>
<td>0.1861</td>
</tr>
<tr>
<td>prime:condition</td>
<td>0.0716</td>
<td>0.9648</td>
</tr>
<tr>
<td>trans</td>
<td>0.57</td>
<td>0.4502</td>
</tr>
<tr>
<td>trans:condition</td>
<td>0.5271</td>
<td>0.7683</td>
</tr>
<tr>
<td>trans:prime</td>
<td>2.2827</td>
<td>0.1308</td>
</tr>
<tr>
<td>pronouns</td>
<td>20.649</td>
<td>&lt; 0.001 ***</td>
</tr>
<tr>
<td>pronouns:prime</td>
<td>17.593</td>
<td>&lt; 0.001 ***</td>
</tr>
<tr>
<td>they_frequency</td>
<td>10.014</td>
<td>0.01845 *</td>
</tr>
<tr>
<td>age</td>
<td>2.4414</td>
<td>0.1182</td>
</tr>
<tr>
<td>age:prime</td>
<td>1.6651</td>
<td>0.1969</td>
</tr>
<tr>
<td>age:condition</td>
<td>3.0748</td>
<td>0.2149</td>
</tr>
</tbody>
</table>

The analysis returned a main effect antecedent condition. Figure 6 breaks down pronoun production by AC, showing that the highest rate of singular they production compared to he and she occurred in the gender-neutral name condition. At 69.75%, this presents almost double the percentage of they-production in the gender-biased name condition at 35.29%. This finding may be interpreted to align with Kramer et al. (2022), who report significantly higher production of singular they in reference to androgynously-presenting people compared to photos depicting stereotypically feminine or masculine individuals. Singular they was produced least often in comparison to binary pronouns in the gender-biased NP condition (25.58%). Interactions of antecedent condition with other demographic variables did not reach statistical significance.
Other than antecedent condition, participants’ own personal pronouns significantly impacted which pronoun they produced in the task. As Figure 7 shows, participants who go by they/them pronouns produced they/them most often at 71.65%, followed by participants who go by multiple different pronouns at 54.10%. Both binary pronoun groups produced singular they in less than a third of all cases, with the she/her-group leading at 28.09% over the he/him-group at 19.97%.
While Kramer et al. (2022) do not report a main effect of participant gender, they do find a significant interaction of gender and priming. Similarly, I find a significant interaction of pronouns and priming. Figure 8 shows the proportion of singular they production for different pronoun groups divided by prime and control group members. Interestingly, while the binary pronoun users show a higher proportion of singular they in the priming group than in the control group, the opposite pattern emerges in the they/them and multiple-pronoun group. However, even the binary pronoun group with the highest singular they percentage (she/her-users in the priming group: 41.44%) lie below the gender-diverse group with the lowest percentage of they-production (multiple-pronoun users in the priming group: 50.00%).
Participants’ self-reported frequency of interacting with they/them-users was also found significant in this analysis. Figure 9 illustrates how the daily interaction group produce the highest percentage of singular they (63.11%), followed in big steps by the weekly (44.44%), monthly (23.81%) and almost never groups (15.32%). This may be compared to Kramer et al. (2022)’s similar measure of familiarity with gender-diverse people, which proved a significant predictor in the written but not oral production task.
Unlike pronouns, trans* identity did not significantly affect production of singular *they*. This contrasts with Hekanaho (2020)’s findings that trans* people were more likely to use non-gendered pronouns, albeit in written production and with epicene referents. Interactions of trans* identity with AC or priming did not reach significance either.

Similarly, there was no significant effect of age, mirroring Kramer et al. (2022). Interactions of age with priming or AC were also not found significant. Both Kramer et al.’s and my own study, however, feature an overwhelmingly younger aged sample.
4.4.1. Excluded Production Data

Finally, a separate analysis was run including participants with a degree-level Linguistics experience and those who had learned English after the age of 5. Anova comparison with the maximum model revealed a significant effect of age at English acquisition, as shown in Table 10. The after-5 group produced singular *they* only 6.22% of the time, whereas *they*-production for the before-5 group lies at 36.10%.

Table 10

*Production Model Comparisons Output with Excluded Participants*

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Chi-squared</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>eng</td>
<td>7.2843</td>
<td>0.006956 **</td>
</tr>
<tr>
<td>eng:condition</td>
<td>1.002</td>
<td>0.6059</td>
</tr>
<tr>
<td>ling</td>
<td>3.2755</td>
<td>0.07032</td>
</tr>
<tr>
<td>ling:condition</td>
<td>5.1648</td>
<td>0.07559</td>
</tr>
</tbody>
</table>
5. Discussion

5.1. Limitations

Before discussing the results of this study, I will consider its limitations and their possible repercussions. First and foremost, the overall number of participants was quite low (n = 65) compared to other studies, which reduces the statistical power of tests. The low number resulted not only from previously discussed exclusion criteria, but also due to paying participants for their time in line with Scottish Living Wage for ethical reasons, which should be considered by researchers especially when collecting data from minority groups, conducting time-consuming experiments and generally in times of economic crises such as the current cost of living crisis in the United Kingdom. The small number of participants consequently meant small numbers of the different levels of predictor variables. Further, the main aim during recruitment was to recruit an even number of cisgender and gender-diverse participants, meaning there was less focus on other demographic variables. The age variable was perhaps most impacted by this, surfacing in a relatively small number of older participants and an overall young and linguistically progressive sample. This may also be a result of both online recruitment and testing, as older people may be underrepresented on online recruitment platforms and social media, as well as within the University environment.

Another issue related to recruitment platforms emerged due to the use of pre-screening criteria on Prolific (www.prolific.co) not being answered correctly by users for different reasons. This was only discovered after users had completed the study, and their subsequent disqualification resulted in loss of data as well as loss of payment for the users, leading to frustration for both parties. The solution was to manually re-iterate Prolific’s pre-screening questions to the beginning of the online study, which had the unwanted side effect of revealing the focus of the study being singular they early on.

This leads to the broader issue of self-reported survey data, which does not necessarily reflect reality. Particularly in survey questions where participants need to estimate, for example how often they interact with people who go by they/them-pronouns, it is hard for the participants themselves to self-report with absolute certainty. Therefore, results must be interpreted with caution.
Further, even outwith Prolific, many participants realised the main variables of interest during the course of the study. In future research, this may be avoided by conducting separate, short experiments rather than one long study hosting several experiments. It may be the case that the results of the reading and production experiment may have differed if they had been situated at the beginning of the study. Especially for the production experiment, it is unclear whether any auditory priming effects may have been influenced by the previous exposure to written singular *they* which may have served as a prime in itself.

Finally, due to singular *they* functioning as a gender-neutral, gender-inclusive as well as personal pronoun, it is hard if not impossible to check which use participants activated when reading or speaking the pronoun. For example, in the production task, some participants may have used *they* to refer to *Wai* due to not knowing the referent’s gender, not having a strong feminine or masculine gender-bias for this particular name in their grammar or remembering that *Wai* was referred to as *they* in the previous short story. Therefore, care must be taken when analysing the results to account for multiple possible interpretations. With these limitations in mind, I now go on to discuss the wider implications of the study results, starting with each experiment separately and concluding with a general discussion.

### 5.2. Naturalness Rating

The naturalness ratings largely patterned in accordance with the three different group/stages of singular *they* users as identified by Bjorkman (2017) and Konnelly and Cowper (2020). I observe a general increase in naturalness from given names and gender-biased NP < specific definite NP < indefinite epicene NP and quantified NP antecedents, and only a few possible exceptions to the implicational hierarchy. This further supports the prior assumption that English and American English language users exhibit similar rating behaviours concerning singular *they* (Hekanaho, 2020). However, predictions of an age effect with younger speakers providing higher ratings (Conrod, 2019) are not borne out, probably due to an unbalanced sample with less older than younger speakers. Thus, no conclusions can be drawn regarding a language change in process as based on age effects.

But what other factors might be driving the singular *they* language change? A main dichotomy may be posited by comparing simple frequency of exposure to a more complex underlying change in the language user’s mental representation of gender. Starting with the latter, I focus on participants’ pronouns and trans* identity as possible clues. In my analysis,
pronouns and trans* identity behave differently from each other. While participants who identify within the trans* umbrella rate singular they significantly more natural than cis participants across all antecedent conditions, participants who go by they/them pronouns provide significantly higher ratings than she/her and he/him participants only in the stage 3 or super-innovative antecedent conditions: gender-biased NPs and given names. Multiple-pronoun users pattern similarly to the they/them-group, except that they do not rate gender-biased given names significantly more natural. Recalling the high variance inflation factor of pronouns, gender and sexuality, I will compare they/them-users and multiple-pronoun users to non-binary participants in other studies, but it should be noted that other genders are included in these two pronoun groups, e.g. genderqueer, genderfluid and agender.

Conrod (2019) reports a main effect of participant gender, as well as a main effect of transgender identity and an interaction of trans identity with antecedent condition. Men and women rate singular they significantly lower than people that report their gender as “neither”; trans people rate the pronoun significantly higher than cis people, but the interaction reveals that they rate the quantified condition less natural than the other conditions. Conrod notes the high overlap of the “neither” gender category with transgender participants, and it may be that the difference in my results is caused by a relatively higher number of gender-diverse participants within the sample. Another possible explanation are the differing Likert scales (Conrod’s has 7 points while this study has 6).

Regardless, these results indicate that language users’ mental concept of gender influences their feelings of naturalness towards singular they. People who use they/them as their own personal pronoun have a more nuanced mental understanding of gender unlike the male-female gender binary, and therefore perceive singular they as natural in reference to antecedents that may evoke strong masculine or feminine gender associations in people who orient to a rigidly binary gender concept: given names as well as proper nouns with high gender-bias such as kinship terms. It is not immediately clear, however, why people who go by at least two different pronouns do not rate singular they significantly more natural than she/her- and he/him-users if it refers to a given name antecedent with strong masculine or feminine gender bias, even though they do rate it significantly higher in the gender-biased NP and gender-neutral name condition. Referring back to the possible different reasons for using multiple pronouns, it could be that this particular sample includes people who are forced to use one particular pronoun in less accepting social environments that differs from the
pronoun they go by in more accepting environments (Callaway, 2022), and it may be that exposure to these environments enforces binary-gendering of gender-biased given names. Or in other words, these environments may tolerate use of epicene/gender-neutral singular they and gender-biased given names may not be viewed as a possible site for gender-neutrality. In the post-experiment survey, all multiple-pronoun users provide high ratings for the gender-biased sentence along with comments that mention the possibility of the referent using they/them-prouns. However, one participant remarks that the referent (“Jack”) “might identify as they or he/they. Saying they makes sure that you can’t get it wrong”, indicating a possible acknowledgement of the masculine gender bias in co-existence with diverse gender identities, as well as a reading of they as gender-neutral. Overall, it must be reiterated that only seven participants within the sample reported using multiple pronouns, so these stipulations should not be given too much weight. Future research may investigate multiple-pronoun use in a more in-depth manner.

Returning to the alternative possible driving factor behind singular they changing, I consider exposure to singular they in the form of self-reported frequency of interacting with people who go by they/them. Results showed that participants who reported daily or weekly interaction rated singular they significantly more natural across all antecedent groups than those who reported fewer interaction. This mirrors findings by Camilliere et al. (2021), who report higher ratings across a wide range of antecedent types by participants who are more familiar with non-binary genders. As Camilliere et al.’s study features predominantly binary gender participants, my study further expands these findings by confirming the same effect in a sample of more balanced gender identities.

Further, I also find a significant interaction of personal pronoun they-exposure and antecedent condition, with the most pronounced differences located in the stage 3 conditions: gender-biased NP, gender-neutral name and gender-biased name antecedent conditions. Focussing on the given name conditions first, the monthly and almost never group rate personal pronoun they considerably lower than daily and weekly groups – in the neutral name condition, the monthly group provides much higher ratings than the almost never group, whereas the differences between these two in the gender-neutral name condition is much smaller. This supports previous findings by Ackerman (2018a) of more experience with gender-nonconformity co-occurring with significantly higher acceptability ratings of singular they in reference to given name antecedents. Comparing the given name conditions to the
gender-biased NP condition, however, I find that the *monthly* group rates singular *they* much lower than the *almost never* group does. While participants who report almost never interacting with people who go by *they/them* rate the three super-innovative conditions similarly low – below the 3.5 mid-point of the naturalness scale – those who report monthly interaction only rate the gender-biased NP condition below the 3.5 mid-point.

I believe this supports Konnelly and Cowper (2020) descriptions of stage 1 and stage 2 users. Konnelly and Cowper posit that gender features are contrastive in both stages, but their assignment is optional for stage 2 users. Drawing upon Ackerman (2019)’s three-tiered schema of encoding gender, participants who report interacting with *they/them* users on a monthly basis are exposed to more exemplars of a given name occurring in reference to singular *they* than the *almost never* group. Based on these exemplars, Konnelly and Cowper hypothesise that stage 2 users may then either remove the contrastive gender feature from a specific name or create a new lexical entry of this name in their grammar that may map to one particular *they/them*-user. Through more frequent exposure, the exemplar tier will feed into and possibly modify the category tier. This explains why, for the *monthly* group, given name conditions are rated higher than the gender-biased NP condition – due to experience, they are used to singular *they* as a possible personal pronoun, but the low ratings for gender-biased NPs show that they still have contrastive gender features and have not reached stage 3 yet.

Concluding this section, I believe the results support previous theories that while frequent exposure to singular *they* can initiate language change regarding perceived naturalness of the pronoun, it is language users’ mental concept of gender that ultimately determines the completion of this change. Nonetheless, rating data only illuminates one side of a complex phenomenon. I now turn to discuss the outcomes of the reading time experiment.

### 5.3. Reading Time

The self-paced reading time (RT) task returned a somewhat unexpected outcome, namely the non-significant main effect of antecedent condition (AC) that has been reported in most previous literature (Ackerman, 2018c; Block, 2019; Doherty & Conklin, 2017; Foertsch & Gernsbacher, 1997; Han & Moulton, 2022). As this main antecedent effect has been reported in studies on British, Canadian and US-American English, the Britishness of this sample does not provide an explanation for the absence of this main effect. Instead, I propose the reason
may lie in the gender diversity and balance of this sample contrasting with previous RT studies, which is supported by the significant interaction between participants’ pronouns and antecedent condition.

Recalling the predictions made on the basis of earlier studies, it was expected that the gender-biased NP condition would be read significantly slower compared to conditions without gender-bias (Foertsch & Gernsbacher, 1997; Han & Moulton, 2022). Further, if Han and Moulton’s theory is correct and all referential gender-biased antecedents incur a processing cost compared to quantified antecedents, slower reading times would also be predicted for the gender-biased name condition. The pronoun-AC interaction reveals that these predictions are only confirmed in participants who use he/him pronouns. While they/them participants also read the gender-biased NP condition much slower than the other ACs, gender-biased names do not show a spike in reading time. In the she/her group, as well as for the multiple-pronoun group, variation in-between conditions is much smaller.

While these results do not necessarily refute Han and Moulton (2022)’s theory, they certainly call into question its universal application. As Han and Moulton’s paper does not feature a demographic participant breakdown, it can only be speculated that their sample may have featured more men than women. Based on my own results, I propose that theories about the underlying features of singular they should at least acknowledge, if not analyse, possible differences due to participant gender. Further, I suggest that if gender-diverse language users are to form a substantial part of the analysis, researchers should strive for a 1:1 ratio of cisgender and gender-diverse participants, as Block (2019) does not report significant RT differences in a sample with an approximate 2:1 ratio.

I believe the differences between the multiple-pronoun group and they/them-group may also warrant further research. If replicated, it may show linguistic diversity within gender-diverse language communities and make a case against collapsing this diversity into another binary of cis vs gender-diverse language users. This case is further supported by the findings showing considerably different patterns in she/her-users compared to he/him-users, as well as the non-significance of trans* identity and its interaction with AC.

Focussing on this diversity in my sample, it is not immediately clear why they/them-users read sentences with singular they referencing a gender-biased NP much slower than other conditions, including gender-biased name referents. Applying the three-stage model of
singular *they* language change, it could be possible that the *they/them*-group features mostly stage 2 users who have encountered enough *they*-name exemplars to not apply contrastive binary gender features to names, but still apply them to gender-biased NP referents. However, it seems counter-intuitive that more people who use *they/them*, but less people who use *she/her*, would have a strongly binary mental representation of gender, especially considering the higher proportion of trans* participants within the *they/them*-group compared to the *she/her*-group. Instead, I consider (Ackerman, 2018c)’s proposal of different underlying reasons for reading delay. It might very well be the case that *they/them*-users read gender-biased NP sentences slower to reaccommodate the new information that singular *they* conveys, namely that a specific referent may use *they/them* pronouns. This personal pronoun reading of *they* may not be available to the majority of the *he/him*- and *she/her*-group in this study, who may be applying a solely gender-neutral/epicene reading to *they*. And as women have been shown to lead language change, *she/her*-users do not show longer RTs for gender-biased NP singular *they* while *he/him*-users may briefly struggle to process a gender-neutral pronoun with a gender-biased referent (both NP and given name).

But why does the pattern for multiple-pronoun users not show slower RTs for gender-biased NP antecedents? Reconsidering the rating patterns in the previous section, it might be that multiple-pronoun users pattern more similar with binary pronoun users if their use of multiple different pronouns is motivated by unaccepting environments. However, it needs reiterated that this is a purely speculative line of reasoning and people’s reasons for using several pronouns are multi-faceted. As this sample only features seven multiple-pronoun users, not too much importance should be placed upon these results and further research is needed to establish their robustness. In particular, ERP research on processing singular *they* by gender-diverse language users may provide answers regarding the underlying reasons for RT effects.

5.4. **Oral Production after Auditory Priming**

Oral production of singular *they* was significantly affected by antecedent condition, with the neutral name condition favouring *they*-production compared to the gender-biased name and NP conditions. This may be compared to a similar pattern emerging in Kramer et al. (2022)’s study where singular *they* is predominantly produced in reference to pictures of androgynously-presenting individuals, both in oral and written production. However, Kramer
et al. also report a positive priming effect for they-use in the androgynous condition in their oral production experiment, while my analysis does not return a significant main effect of priming nor a significant interaction of prime and antecedent condition.

In fact, my results are more similar to Kramer et al. (2022)’s written production task than the oral task, as I find a significant interaction between pronouns/gender and priming as well as a main effect of self-reported familiarity/frequency of exposure to singular they as a personal pronoun. As for the pronoun-prime interaction, like Kramer et al., I find that he/him-users (men) in the priming group are less likely to use singular they than other pronoun/gender priming groups; nevertheless, both he/him- and she/her-users are more likely to produce singular they in the priming group than in the control group. The relatively small priming effect shows that while simply using they/them pronouns may reduce the chance of misgendering, it is not effective on its own and more explicit measures are likely needed, such as name tags and pronoun statements (Gardner & Brown-Schmidt, 2023), as well as correcting people if they misgender someone. In addition, I also find what appears to be a reverse priming pattern for they/them-users, and to a smaller effect in multiple-pronoun users, as these groups produce singular they at a much higher rate in the control group than in the priming group. It is not immediately clear why such a pattern might arise, but unless replicated and found significant by future research, it should not be given too much importance.

However, I also report a significant main effect of participant’s pronouns: regardless of priming, the they/them- and multiple-pronoun group produce singular they considerably more often than she/her- and he/him-users. Kramer et al. (2022) do not report a significant main effect of gender, and my results suggest that this is likely due to their strongly binary sample (2 non-binary participants out of 115). This is further supported by Hekanaho (2020)’s findings of trans* participants being significantly more likely to produce singular they with generic antecedents, noting that Hekanaho collapsed non-binary participants and transgender participants that are not non-binary into a “transgender” category, due to the relatively low number of “binary” trans participants.

The significant main effect of frequent interaction with they/them-users again mirrors Kramer et al. (2022)’s effect of familiarity in the writing task. The order of decreasing production by interaction in the daily > weekly > monthly > almost never groups does intuitively make sense, but it must be noted that the daily group does not produce singular
they at ceiling, and the almost never group does not always produce he or she pronouns in favour of them. Unfortunately, modelling did not allow inclusion of any interactions of they-frequency with other variables due to convergence issues, but hopefully future research may explore this in more detail.

To sum up the discussion of production results, I found multiple similarities with US-American studies, suggesting similar production patterns of singular they in US-American and British English. Similarities with Kramer et al. (2022) are found predominantly in their writing task compared to their oral production task; however, I propose this may be due to the availability of the “prime” throughout Kramer et al.’s production task compared to the priming before production in their writing as well as my own oral production experiment. I believe this is further supported by Gardner and Brown-Schmidt (2021) finding that singular they priming does not significantly increase production in an experiment which features distractor questions. Instead, Gardner and Brown-Schmidt show that both memory and production of singular they is not effected by priming on its own. I further argue that inclusion, and differentiation, of gender-diverse participants can reveal important demographic effects that are lost in binary cisgender samples.

5.5. General Discussion

Reviewing the predictions made on previous literature, the general expectation of demographic data impacting rating and production but not reading time data has been broadly confirmed. Naturalness ratings and oral production showed significant influence of participant’s gender (be it in the form of pronouns or trans* identity) or their exposure to other people who use singular they as their own pronoun. Against expectations, however, the language-internal factor of antecedent condition on its own only emerged as a significant predictor in the oral production task. Keeping in mind that the production experiment featured three out of the six antecedent conditions tested in rating and reading, it seems probable that this main effect would not have appeared if all six AC were tested in production. This proposition is supported by the patterns emerging in the significant interactions of AC with language-external variables in rating and RT data, as the three “super-innovative” conditions (gender-biased NP and names, gender-neutral names) tend to show the most variation in-between levels.
Both rating and production data show (differing) main effects or interactions that include participant’s own gender as well their exposure to other people who use singular *they*. Comparing this to the lack of a *they*-frequency effect in the RT task, a possible deduction may be that while rating and production may be influenced simply through frequency of linguistic exposure to singular *they*, reading time and therefore processing are impacted by an individual’s mental concept of gender. However, the RT patterns of different pronoun groups do not correspond to a straightforward mapping of *they/them* pronouns to fast RTs and binary pronouns to slower RTs as a more nuanced picture emerges. Until more processing, and crucially, ERP studies, employ a trans linguistic approach to the study of singular *they*, the underlying reasons for the observed patterns remain opaque.

Based on these observations, it becomes clear how sociolinguistic analysis, and in particular from a trans linguistics angle, is crucial when researching singular *they*, as it adds a nuanced understanding to the interaction of language-internal and -external factors. I believe the most striking point for this argument can be made using the RT data analysis, which reveals the widely reported effect of antecedent condition on singular *they* reading time differs based on which pronouns participants go by. Recalling Block (2019)’s thesis comparing naturalness rating and reading of singular *they* between cisgender and non-binary participants, Block finds that participant gender only impacts rating but not reading time. While Block’s experimental design differs in various ways from my own, both in materials and procedures, I believe the decisive factor resulting in diverging outcomes may be attributed to the ratio of cisgender to non-binary participants. Moreover, I find differences between participants who only use *they/them* pronouns and those who go by multiple different pronouns, e.g. *they* and *she*, while Block’s analysis combines these into one non-binary category.

However, the multiple significant interactions of antecedent condition with demographic variables also underlines the importance of language-internal factors governing perception and production of singular *they*. While membership of the three groups or stages of singular *they* “acceptance” is likely influenced by demographic variables such as age and gender, membership criteria is based on the rating of singular *they* with different antecedent types. Due to the use of judgement/rating data by Bjorkman (2017), Konnelly and Cowper (2020) and others to establish the three groups/stages, as well as the recognition of the inherent link between demographic characteristics and group membership, most rating studies on singular
they use the established framework whose results are mostly confirmed in my study. Comparatively, experimental production studies on singular they are also adapting these trans linguistics frameworks, whereas processing studies seem to lag behind, often not even reporting demographic participant make-up, much less analysing it. This is especially dangerous if reading time data is posited as a one-to-one mapping to processing and thus “implicit” grammar as a contrast to “explicit” rating data. Using outdated terminology such as “number-mismatch” and referring to they as a plural pronoun further compounds this problem. Singular they is inextricably linked to gender-diverse communities, and researchers must acknowledge these implications regardless of which discipline they are from. I believe my study has shown that no theory of the underlying grammar of singular they can be proposed without acknowledging the diverse patterns of perception and production in different communities of language users.
6. Conclusion

Having discussed the results of my experiments, I will now revisit the three questions posed at the start of my research:

1. Which language-internal and -external factors influence perception and production of singular they? Is the current language change being driven by language user’s mental concept of gender, or is it simply a question of frequent exposure to singular they?
2. Do rating, reading and production data show the same patterns?
3. Does a trans linguistic approach to the study of singular they lead to new insights?

To address the first two points, I have found significant main effects and interactions of both the language-internal factor of antecedent condition as well as various language-external factors on rating, reading and production of singular they. Reading time data showed a significant interaction of pronouns and antecedent condition only. Oral production, on the other hand, was significantly influenced by antecedent condition, self-reported exposure to they/them pronouns, participant’s own pronouns as well the interaction of pronouns with auditory priming. Finally, naturalness rating showed significant effects of interactions of antecedent condition with pronouns as well as with self-reported frequency of they/them exposure, the latter of which also showed a significant main effect. Other main effects on naturalness ratings were participants’ trans* identity as well as whether they rated a sentence as part of their own speech compared to someone else’s.

Overall, these results show asymmetries between the three data types, which is not dissimilar to findings reported in previous literature. However, the nature of these results revealed some novel insights which leads to the third question. Namely, the significant interaction of pronouns with antecedent condition had not been reported in reading time studies before; instead, studies had found no significant effects of any demographic variables with antecedent condition being the only significant factor. Further, I also report different effects of participant’s pronouns compared to their trans* identity, which emphasizes the importance of nuanced data collection of participant genders. Regarding the effect of auditory priming, I report small priming effects in binary pronoun users, suggesting that simply using correct pronouns may limit but not eliminate instances of misgendering, and more pro-active measures are likely needed.
In conclusion, I hope that I have shown that a trans linguistics approach is not only beneficial but, in fact, crucial when researching and reporting on singular *they*. In addition to the added scientific value that comes with acknowledging linguistic diversity, it is more important than ever to acknowledge the autonomy of gender-diverse people – not only regarding linguistic features but also as core demographics in data and in the real world. Not every study will be able to recruit a high number of gender-diverse participants, but every researcher chooses how to frame their own data in regard to the wider population. Studies on singular *they*, particularly those seeking to explain underlying grammatical features, should be clear on the data that informs their theories – while not every study can feature a balanced gender-diverse sample, it should at least be acknowledged if this is the case. Especially when examining language indexing minority group membership, researchers must take care to question existing normative assumptions that are based on majority models.
7. References

Ackerman, L. (2018a). *Being themself: processing and resolution of singular (im)personal they* [Conference presentation abstract] CUNY, University of California, Davis. https://osf.io/zmpjc/

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

Ackerman, L. (2018c). *Processing singular they with generic and specific antecedents* [Conference poster presentation]. Architectures and Mechanisms for Language Processing (AMLaP) Berlin. https://drive.google.com/file/d/1YnJXtWLOt3uXTh4WZLokE7RYWjQ2IYes/view

Ackerman, L. (2018d). Toward an understanding of how nonbinary pronouns are learned and processed [Conference poster presentation]. ZAS Leibniz, Berlin. https://drive.google.com/file/d/1Q9RJFynALE5Dq3CVeurbm08ZfOy9tLUn/view


Ackerman, L., Riches, N., & Wallenberg, J. (2018, 05/01/2018). Coreference dependency formation is modulated by experience with variation of human gender The 92nd Annual Meeting of the Linguistic Society of America (LSA2018), Salt Lake City, UT. https://drive.google.com/file/d/1yEl1eJReIU-c7u3y7kBgyUWkHEvTwdc7/view


Farrington, B. (2023). DeSantis signs bills targeting drag shows, transgender kids and the use of bathrooms and pronouns. _AP News_. Retrieved 30 May 2023, from


Farrington, B. (2023). DeSantis signs bills targeting drag shows, transgender kids and the use of bathrooms and pronouns. _AP News_. Retrieved 30 May 2023, from


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


https://doi.org/10.1037/0003-066X.33.11.1032

https://doi.org/10.1017/S0047404500015529


https://doi.org/10.31234/osf.io/hwzke


https://doi.org/https://doi.org/10.1016/bs.pbr.2020.06.005


Linguistics. In K. Hall & R. Barrett (Eds.), The Oxford Handbook of Language and
Sexuality Oxford University Press.
https://doi.org/10.1093/oxfordhb/9780190212926.013.45
## 8. Appendix

### 8.1. Appendix A: Norming Study

<table>
<thead>
<tr>
<th>Expected gender association</th>
<th>Name</th>
<th>Mean rating</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>masculine</td>
<td>Robert</td>
<td>1.1</td>
<td>0.402578</td>
</tr>
<tr>
<td>masculine</td>
<td>David</td>
<td>1.233333</td>
<td>0.430183</td>
</tr>
<tr>
<td>masculine</td>
<td>Andrew</td>
<td>1.266667</td>
<td>0.639684</td>
</tr>
<tr>
<td>masculine</td>
<td>Jack</td>
<td>1.666667</td>
<td>0.958927</td>
</tr>
<tr>
<td>masculine</td>
<td>James</td>
<td>1.766667</td>
<td>1.222866</td>
</tr>
<tr>
<td>masculine</td>
<td>Aiden</td>
<td>2</td>
<td>1.144703</td>
</tr>
<tr>
<td>non-binary</td>
<td>Ashton</td>
<td>2.2</td>
<td>1.270352</td>
</tr>
<tr>
<td>masculine</td>
<td>Hunter</td>
<td>2.233333</td>
<td>1.381736</td>
</tr>
<tr>
<td>non-binary</td>
<td>Cody</td>
<td>2.333333</td>
<td>1.093345</td>
</tr>
<tr>
<td>masculine</td>
<td>Arlo</td>
<td>2.833333</td>
<td>1.341212</td>
</tr>
<tr>
<td>non-binary</td>
<td>Ezra</td>
<td>2.966667</td>
<td>1.519604</td>
</tr>
<tr>
<td>non-binary</td>
<td>Rowan</td>
<td>2.966667</td>
<td>1.376736</td>
</tr>
<tr>
<td>masculine</td>
<td>Jordan</td>
<td>3.233333</td>
<td>1.104328</td>
</tr>
<tr>
<td>non-binary</td>
<td>Beau</td>
<td>3.4</td>
<td>1.476249</td>
</tr>
<tr>
<td>non-binary</td>
<td>Marley</td>
<td>3.466667</td>
<td>1.252125</td>
</tr>
<tr>
<td>non-binary</td>
<td>Harley</td>
<td>3.666667</td>
<td>1.154701</td>
</tr>
<tr>
<td>non-binary</td>
<td>Alex</td>
<td>3.7</td>
<td>0.83666</td>
</tr>
<tr>
<td>non-binary</td>
<td>Quinn</td>
<td>3.766667</td>
<td>1.356551</td>
</tr>
<tr>
<td>non-binary</td>
<td>Regan</td>
<td>3.8</td>
<td>1.494819</td>
</tr>
<tr>
<td>non-binary</td>
<td>Jaimie</td>
<td>3.933333</td>
<td>1.412587</td>
</tr>
<tr>
<td>non-binary</td>
<td>Wai</td>
<td>3.966667</td>
<td>0.556053</td>
</tr>
<tr>
<td>non-binary</td>
<td>River</td>
<td>4</td>
<td>1.286535</td>
</tr>
<tr>
<td>non-binary</td>
<td>Frankie</td>
<td>4.033333</td>
<td>1.098065</td>
</tr>
<tr>
<td>masculine</td>
<td>Riley</td>
<td>4.1</td>
<td>1.155198</td>
</tr>
<tr>
<td>non-binary</td>
<td>Taylor</td>
<td>4.133333</td>
<td>0.681445</td>
</tr>
<tr>
<td>non-binary</td>
<td>Shay</td>
<td>4.3</td>
<td>1.178836</td>
</tr>
<tr>
<td>non-binary</td>
<td>Casey</td>
<td>4.733333</td>
<td>1.172481</td>
</tr>
<tr>
<td>feminine</td>
<td>Robyn</td>
<td>4.766667</td>
<td>1.194335</td>
</tr>
<tr>
<td>Gender</td>
<td>Name</td>
<td>Rating 1</td>
<td>Rating 2</td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>non-binary</td>
<td>Blair</td>
<td>4.766667</td>
<td>1.250747</td>
</tr>
<tr>
<td>non-binary</td>
<td>Kenzie</td>
<td>5.233333</td>
<td>1.135124</td>
</tr>
<tr>
<td>non-binary</td>
<td>Angel</td>
<td>5.366667</td>
<td>1.376736</td>
</tr>
<tr>
<td>feminine</td>
<td>Brooke</td>
<td>5.466667</td>
<td>1.166585</td>
</tr>
<tr>
<td>feminine</td>
<td>Isla</td>
<td>6.3</td>
<td>0.749713</td>
</tr>
<tr>
<td>feminine</td>
<td>Millie</td>
<td>6.533333</td>
<td>0.681445</td>
</tr>
<tr>
<td>feminine</td>
<td>Megan</td>
<td>6.7</td>
<td>0.534983</td>
</tr>
<tr>
<td>feminine</td>
<td>Emily</td>
<td>6.733333</td>
<td>0.520831</td>
</tr>
<tr>
<td>feminine</td>
<td>Elizabeth</td>
<td>6.766667</td>
<td>0.678911</td>
</tr>
<tr>
<td>feminine</td>
<td>Olivia</td>
<td>6.766667</td>
<td>0.504007</td>
</tr>
<tr>
<td>feminine</td>
<td>Mary</td>
<td>6.833333</td>
<td>0.461133</td>
</tr>
<tr>
<td>feminine</td>
<td>Sophie</td>
<td>6.866667</td>
<td>0.434173</td>
</tr>
</tbody>
</table>

### 8.2. Appendix B: Naturalness Rating Task

Set 1:

**Item 1:** Every jogger likes to stretch their legs after running. (Condition I)
A responsible jogger likes to stretch their legs after running. (Condition II)
This jogger likes to stretch their legs after running. (Condition III)
The boy likes to stretch their legs after running. (Condition IV)
Taylor likes to stretch their legs after running. (Condition V)
Sophie likes to stretch their legs after running. (Condition VI)

**Item 2:** Every musician needs to play their own songs.
An aspiring musician needs to play their own songs.
This musician needs to play their own songs.
The boy needs to play their own songs.
Taylor needs to play their own songs.
Sophie needs to play their own songs.

**Item 3:** Every customer wants to get their money back.
An unhappy customer wants to get their money back.
This customer wants to get their money back.
The boy wants to get their money back.
Taylor wants to get their money back.
Sophie wants to get their money back.

Item 4: Every cyclist ought to look over their shoulder before turning.
A good cyclist ought to look over their shoulder before turning.
This cyclist ought to look over their shoulder before turning.
The boy ought to look over their shoulder before turning.
Taylor ought to look over their shoulder before turning.
Sophie ought to look over their shoulder before turning.

Item 5: Every baby loves to see their favourite animal.
A curious baby loves to see their favourite animal.
This baby loves to see their favourite animal.
The boy loves to see their favourite animal.
Taylor loves to see their favourite animal.
Sophie loves to see their favourite animal.

Item 6: Every child has to finish their breakfast before going outside.
A growing child has to finish their breakfast before going outside.
This child has to finish their breakfast before going outside.
The boy has to finish their breakfast before going outside.
Taylor has to finish their breakfast before going outside.
Sophie has to finish their breakfast before going outside.

Fillers

Ungrammatical:
Reading never she likes be.
And Bill grandpa some very no again saw.

Non-standard:
You be lookin' at them people.
I go swimming anymore.
Wren is wanting to go out.
We’ll can try again later.
Standard:

The drivers here are getting on my nerves.
Unfortunately, the children already ate all the chocolate.
Lately, the radio has been playing a lot of classical music.
Most people brush their teeth after breakfast.
That’s not possible.
Please don’t smoke around the children.

8.3. Appendix C: Reading Time Task
Set 1: see Appendix B

Set 2: Item 7: Every teenager will disagree with their own parents.
   A rebellious teenager will disagree with their own parents.
   This teenager will disagree with their own parents.
   The woman will disagree with their own parents.
   Frankie will disagree with their own parents.
   Mary will disagree with their own parents.

Item 8: Every cook should keep their own kitchen clean.
   A dutiful cook should keep their own kitchen clean.
   This cook should keep their own kitchen clean.
   The woman should keep their own kitchen clean.
   Frankie should keep their own kitchen clean.
   Mary should keep their own kitchen clean.

Item 9: Every author thinks about self-publishing their book.
   A rich author thinks about self-publishing their book.
   This author thinks about self-publishing their book.
   The woman thinks about self-publishing their book.
   Frankie thinks about self-publishing their book.
   Mary thinks about self-publishing their book.

Item 10: Every lawyer can park their car for free behind the court house.
A working lawyer can park their car for free behind the court house.
This lawyer can park their car for free behind the court house.
The woman can park their car for free behind the court house.
Frankie can park their car for free behind the court house.
Mary can park their car for free behind the court house.

Item 11: Every person could donate their kidney if needed.
A healthy person could donate their kidney if needed.
This person could donate their kidney if needed.
The woman could donate their kidney if needed.
Frankie could donate their kidney if needed.
Mary could donate their kidney if needed.

Item 12: Every hotel guest should return their keys before leaving.
A polite hotel guest should return their keys before leaving.
This hotel guest should return their keys before leaving.
The woman should return their keys before leaving.
Frankie should return their keys before leaving.
Mary should return their keys before leaving.

Set 3: Item 13: Every patient takes their medicine in order to recover.
A sick patient takes their medicine in order to recover.
This patient takes their medicine in order to recover.
The girl takes their medicine in order to recover.
Alex takes their medicine in order to recover.
Robert takes their medicine in order to recover.

Item 14: Every fan hopes to meet their hero before dying.
An old fan hopes to meet their hero before dying.
This fan hopes to meet their hero before dying.
The girl hopes to meet their hero before dying.
Alex hopes to meet their hero before dying.
Robert hopes to meet their hero before dying.
Item 15: Every celebrity hides their face from the paparazzi.
   A tired celebrity hides their face from the paparazzi.
   This celebrity hides their face from the paparazzi.
   The girl hides their face from the paparazzi.
   Alex hides their face from the paparazzi.
   Robert hides their face from the paparazzi.

Item 16: Every spectator cranes their neck to see what’s happening.
   A short spectator cranes their neck to see what’s happening.
   This spectator cranes their neck to see what’s happening.
   The girl cranes their neck to see what’s happening.
   Alex cranes their neck to see what’s happening.
   Robert cranes their neck to see what’s happening.

Item 17: Every doctor tries their best every day.
   A great doctor tries their best every day.
   This doctor tries their best every day.
   The girl tries their best every day.
   Alex tries their best every day.
   Robert tries their best every day.

Item 18: Every photographer takes their own camera to work.
   A free-lance photographer takes their own camera to work.
   This photographer takes their own camera to work.
   The girl takes their own camera to work.
   Alex takes their own camera to work.
   Robert takes their own camera to work.

Set 4: Item 19: Every parent has stubbed their toe on a child’s forgotten toy.
   An exhausted parent has stubbed their toe on a child’s forgotten toy.
   This parent has stubbed their toe on a child’s forgotten toy.
   The man has stubbed their toe on a child’s forgotten toy.
   Harley has stubbed their toe on a child’s forgotten toy.
   Jack has stubbed their toe on a child’s forgotten toy.
Item 20: Every student enjoys writing their dissertation.
   A fortunate student enjoys writing their dissertation.
   This student enjoys writing their dissertation.
   The man enjoys writing their dissertation.
   Harley enjoys writing their dissertation.
   Jack enjoys writing their dissertation.

Item 21: Every poet hates listening to their own voice.
   An insecure poet hates listening to their own voice.
   This poet hates listening to their own voice.
   The man hates listening to their own voice.
   Harley hates listening to their own voice.
   Jack hates listening to their own voice.

Item 22: Every pedestrian fears for their safety when crossing multi-lane roundabouts.
   A cautious pedestrian fears for their safety when crossing multi-lane roundabouts.
   This pedestrian fears for their safety when crossing multi-lane roundabouts.
   The man fears for their safety when crossing multi-lane roundabouts.
   Harley fears for their safety when crossing multi-lane roundabouts.
   Jack fears for their safety when crossing multi-lane roundabouts.

Item 23: Every entertainer loves seeing their audience laugh at a joke.
   A confident entertainer loves seeing their audience laugh at a joke.
   This entertainer loves seeing their audience laugh at a joke.
   The man loves seeing their audience laugh at a joke.
   Harley loves seeing their audience laugh at a joke.
   Jack loves seeing their audience laugh at a joke.

Item 24: Every kid avoids brushing their teeth.
   A lazy kid avoids brushing their teeth.
   This kid avoids brushing their teeth.
   The man avoids brushing their teeth.
Harley avoids brushing their teeth.
Jack avoids brushing their teeth.

**Fillers:**

**Ungrammatical:**

The man the shoe on put.
Reading always she likes be.
Tree within man tall looked all over.
And Bill grandpa some very no again saw.
Shimmer did lake so sea.
Have although rabbits greens.
Do to not or he.
I me and opinions filing goes.

**Non-standard:**

Yous two have to get your act together.
Callum has went to school.
The dug needs fetched from outside
Maryam don't know her neighbour.
They stays at home most days.
Wren is wanting to go home.
There's not a day she hasn't came over.
We'll can try again later.
His father die last year.
They've brung it home with them.
My mother drives slow.
You be lookin' at them people.
The clothes is made from wool.
You've known him for twenty year.
I go swimming anymore.
Tam's no leaving just yet.

**Standard:**
Every day, it gets a little bit warmer.
The cyclists here are getting on my nerves.
Sasha dislikes going on long walks.
Luckily, the children already ate all the chocolate.
A good instructor never loses control.
Lately, the radio has been playing a lot of disco music.
Riding your bike to work every day could be good for you.
Most people brush their teeth before breakfast.
The cat is sitting in the window and watching.
That's not possible.
Staff are advised to finish on time.
Please don't smoke around the children.
This cookbook is useless.
Those spiders are a real problem.
The tide is coming in fast.
Here's your new cardigan.
To complete the test, please keep going.
Benjamina hates assembling furniture.
Your parcel will be delivered tomorrow.
Red goes well with yellow.
Our manager hasn't given up.
None of my friends want to come.
Thirty pounds wouldn't be a bargain.
Nails must be trimmed regularly.

8.4. Appendix D: Production Task

8.4.1. Phase 1: Priming/Control items

Story 1: David is looking for a birthday present for Andrew. Even though David has known [them/him] for five years, no suitable present comes to mind. Andrew likes reading, but David can’t remember whether [they/he] like(s) fiction or non-fiction. After hours of deliberation, the shopping cart remains empty. In the end, David goes home not having bought anything.
Story 2: Our queen has an eldest daughter who loves horse-riding and sword-fighting. The queen asks the princess to decide whether [they/she] want to become queen [themselves/herself]. Not having had a choice, the queen had to take over the throne in younger years. Upon a short reflection, the princess decides that [they/she] do(es) not want to reign and rather serve the kingdom by becoming the first female knight. The queen is not too happy with this answer.

Story 3: Riley has been working together with Wai for two weeks. Wai [themselves/himself] hired Riley but has recently expressed [their/his] dissatisfaction at the hiring choice. Allegedly, Riley has been late too many times and some customers have made complaints. Riley takes note and starts arriving early to every shift, which makes Wai change [their/his] mind. After a few weeks, Riley passes the probation period.

Story 4: My brother’s boyfriend is in trouble. While my brother comes along to every family function, [their/his] boyfriend did not attend the last two weddings, even though my brother said [they/he] would have appreciated the company. The two of them have have been arguing for a while now. While my brother values [themselves/himself] too much to stay in a relationship that does not make [them/him] happy, [their/his] boyfriend is willing to stay together at all costs. It seems that the relationship will be history soon.

Story 5: Megan is offering a lift to Emily, who lives next door. Emily needs to get [their/her] teeth checked and Megan drives past the dentist every day on the way to work. However, Megan only moved in recently and Emily does not feel comfortable accepting the offer. Emily lies about [their/her] tooth ache and walks to the dentist [themselves/herself], but Megan sees [them/her] entering the dentist. The next time they see each other, Megan reassures Emily that all is well.

8.4.2. Phase 1: Filler items
Filler 1: Learning how to play a string instrument is very hard. Not only do you need to master the fine-motor skills required to play, but you also need to have a musical ear. It can take years until you will be able to play a beautiful melody. However, once you do master the instrument, it is one of the sweetest sounds in the world. It just depends on how much time and work you are willing to dedicate.
Filler 2: The forest is an amazing ecosystem! Apart from producing oxygen, it also provides a home for countless animals such as birds, deer, and foxes. Trees can grow incredibly tall and their roots protect the ground from erosion. Moreover, they are beautiful to look at both from below and above. It is such a shame that forests are being decimated because of capitalism.

Filler 3: My mum’s cats are finally starting to get along. Previously, Chichi used to steal Mishi’s food while hissing and growling. Now, however, they have started sitting next to each other. I think I even saw them grooming each other yesterday, but as soon as I came in, Chichi ran away. Hopefully they will become friends at some point.

Filler 4: My grandparents are going to Europe for the first time next week. My grandma is very excited and has already started packing. My grandpa, on the other hand, keeps trying to cancel the flight last minute. Grandpa says their dog needs them, even though the dog is used to staying at my uncle’s house. I’m not sure they’ll make it to the airport in the end!

Filler 5: Our town has a postman who brings a dog to the job. Apparently, the postman used to get chased by dogs daily and decided to get a canine bodyguard. Now, the dogs distract each other and the postman can deliver the post in relative peace. However, the dog is very popular and children come out to play with the dog when the postman comes by. The journey still takes a long time, but now it’s more enjoyable for everyone.

8.4.3. Phase 2: Production

Training Items

Experts say: "We are ready to launch the product."
The cat says: "I'm the best pet of all."
My teacher says: "You are not paying attention."

Critical Items

David says: “I should check the bookshelf.”
Andrew says: “My favourite genre is science fiction.”
The queen says: “I will ask again next year.”
The daughter says: “I am confident in my decision.”
Riley says: “I will never be late again”.
Wai says: “My employee made huge improvements.”
My brother’s boyfriend says: “This was my mistake”
My brother says: “My boyfriend should support me”.
Megan says: “My neighbour has been acting strange lately.”
Emily says: “I need to stop grinding my teeth”.

Fillers
The ombudsman says: “The matter is closed.”
Soledad says: “The sun has been hiding”.
None of the children say: “We want to play first and eat later.”
Fatou and Mamadou say: “Our car keeps breaking down.”
The librarian says: “This book is currently not available.”
Hermine says: “I wish I hadn’t left my phone at home.”
No doctor says: “Eating an apple a day does guarantee health”.
Hatuey says: “I don’t like watching sports anymore.”
Some cyclists say: “Cycling indoors is just not the same.”
My grandma says: “You can’t bake a cake without sugar.”
Ibrahim says: “We don’t need to buy the new console.”
The accountant says: “It has not been a good year for us.”
Jinghua says: “I’m sorry I broke your favourite mug.”
Someone in the back says: “You’re not funny.”
Dumi says: “The flat needs hoovered more often.”
My boss says: “The holidays are so close.”
Chandan says: “Tomorrow is my dog’s birthday.”
The receptionist says: “Unfortunately we are fully booked.”
Lei says: “You forgot to put on sun lotion.”
Our parents say: “We are not going on holiday with you again.”