

Processing bound-variable singular they

CHUNG-HYE HAN

Simon Fraser University, Burnaby, British Columbia, Canada

chunghye@sfu.ca

and

KEIR MOULTON

University of Toronto, Toronto, Ontario, Canada

keir.moulton@utoronto.ca

Abstract

The pronouns *they/them/their* are readily available with a singular interpretation as bound variables (Balhorn 2004, Bjorkman 2017). Referential interpretations are possible, but subject to pragmatic considerations and changes in progress (Bjorkman 2017, Conrod 2019, Konnelly and Cowper 2020). In a series of experiments, we tested differences between bound and referential singular *they* in acceptability and incremental processing, asking whether bound *they* is sensitive to the gender of its antecedent, as referential *they* is (Doherty and Conklin 2017, Ackerman 2018, Ackerman et al. 2018, Conrod 2019). We found that bound singular *they* has an advantage over referential singular *they* in acceptability, even when the antecedent is gendered. In processing, however, bound-variable singular *they* showed a reading time advantage over referential singular *they* only with gendered antecedents. We evaluate these results against existing formal linguistic theories of singular *they* implemented within psycholinguistic models of pronoun processing. We submit that none of the theories fully captures the range of evidence we uncover, in particular the interaction between gender and quantification. We suggest a formal account that does: we propose, using representations from Kratzer (2009) and Sudo (2012), that gender and number features are differentially represented in referential versus binding dependencies. We speculate how this representational difference relates to the

We are extremely indebted to the three anonymous referees for their insightful comments, which were crucial in improving this article. We also thank the members of the Experimental Syntax Lab at Simon Fraser University for their assistance in running the studies reported here. This work was supported in part by Social Sciences and Humanities Research Council (SSHRC) Insight Grant 435-2014-0161 and Academy of Korean Studies Grant AKS-2016-LAB-2250004 to Han, as well as SSHRC Insight Grant 435-2018-1012 to Moulton.

processing mechanisms of antecedent retrieval and to the limited processing advantage for bound singular *they* that we found.

Keywords: bound singular *they*, referential singular *they*, binder index, gender, number

Résumé

Les pronoms anglais *they/them/their* sont facilement disponibles avec une interprétation singulière en tant que variables liées (Balhorn 2004, Bjorkman 2017). Des interprétations référentielles sont possibles, mais dépendent de facteurs pragmatiques et de changements linguistiques en cours (Bjorkman 2017, Conrod 2019, Kannelly et Cowper 2020). Dans une série d'expériences, nous avons examiné les différences entre les *they* singuliers liés et les *they* singuliers référentiels en termes d'acceptabilité et de traitement incrémental, en demandant si le *they* lié est sensible au genre de son antécédent comme l'est le *they* référentiel (Doherty et Conklin 2017, Ackerman et al. 2018, Ackerman et al. 2018, Conrod 2019). Nous avons constaté que le *they* singulier lié a un avantage sur le *they* singulier référentiel en termes d'acceptabilité, même lorsque l'antécédent est genré. Lors du traitement, cependant, la variable liée *they* singulier a montré un avantage en temps de lecture par rapport au *they* singulier référentiel, mais uniquement avec des antécédents genrés. Nous évaluons ces résultats par rapport aux analyses existantes de *they* singulier mises en oeuvre dans les modèles psycholinguistiques de traitement des pronoms. Nous soutenons qu'aucune de ces analyses ne saisit pleinement l'éventail des résultats que nous découvrons, en particulier l'interaction entre le genre et la quantification. Nous suggérons une analyse formelle qui le fait : nous proposons, en utilisant les représentations de Kratzer (2009) et Sudo (2012), que les traits de genre et de nombre sont représentés différemment dans les dépendances référentielles par rapport aux dépendances de liage. Nous spéculons que cette différence de représentation est liée aux mécanismes de traitement, en particulier ceux de la récupération des antécédents, et à l'avantage limité du traitement de *they* singulier que nous avons trouvés.

Mots-clés: *they* singulier lié, *they* singulier référentiel, indice de liage, genre, nombre

1. INTRODUCTION

The experimental, historical, and theoretical literatures have identified a range of properties that allow a morphosyntactically singular noun phrase to antecede the pronouns *they*, *them* and *their*, what we refer to henceforth as *singular they*.¹ A recurring intuition expressed in much of this literature is that bound-variable singular *they* as in (1a) is more acceptable than a referential use (1b) for many speakers.

- (1) a. Every lawyer made *their* case successfully.
- b. The lawyer made *their* case successfully.

This intuition has been confirmed by the most comprehensive experimental work to date on singular *they*, a large-scale acceptability judgment study by Conrod (2019). Conrod asked participants (N=754) to rate singular *they* with different antecedent

¹Abbreviations used: ERP: event-related potential; QP: quantifier phrase; QUANT: quantifier subject; QR: quantifier raising; REF: referential subject; RRT: residual reading time; RT: reading time; SPR: self-paced reading.

types (proper noun, generic and quantified) and collected several participant variables (age, gender, and transgender identity). Conrod found evidence of a change in progress: younger participants gave higher ratings to singular *they* with referential, proper-noun antecedents than older participants, whereas there was no such age effect for the acceptability of singular *they* with generic or quantified antecedents. Conrod (2019) additionally found several complex interactions between age and gender, with non-binary and transgender participants generally rating referential *they* higher. A similar general advantage for bound-variable singular *they* has been found by Camilliere et al. (2019). These studies illustrate a clear asymmetry: bound-variable singular *they* is widely rated as acceptable across speakers, while referential singular *they* shows greater variation.

Where things become yet more interesting is the interaction between gender and quantification. Not only is bound-variable singular *they* highly acceptable to all speakers, there is some evidence that this holds even with gendered antecedents, unlike referential singular *they*. The historical record contains attestations of singular *they* bound by gendered antecedents like *man* (2a) and *sister* (2b).²

- (2) a. There's not a man I meet but doth salute me/As if I were their well-acquainted friend (Shakespeare, *A Comedy of Errors*, 1623)
- b. Both sisters were uncomfortable enough. Each felt for the other, and of course for themselves[.] (Austen, *Pride and Prejudice*, 1813)

In the theoretical syntax literature (Bjorkman 2017, Konnelly and Cowper 2020), such cases are taken to be on par with non-gendered quantified singular *they* – that is, as grammatical for all English users.

We set out to investigate the interplay between antecedent gender and whether singular *they* is bound or referential, examining the consequences for the linguistic representation of singular *they* and for theories of the incremental processing of pronouns. We found that bound-variable singular *they* is indeed rated as more acceptable than referential singular *they*, even with gendered antecedents. However, in two self-paced reading studies we found that these differences in acceptability did not entirely

²Nouns such as *man*, *sister* and *cowgirl* are variously described in the literature as “definitionally gendered” (Kreiner et al. 2008, Ackerman 2019: 8), “lexically gendered” or “gender specific” (Konnelly and Cowper 2020). In this article we describe both these nouns and gender-stereotyped nouns (like *football player*) simply as “gendered” nouns. Nouns that are associated with few or no gender stereotypes (such as *runner*, *student*) we call non-gendered. We use the term “gender” in the sense of “conceptual gender”, as described in Ackerman (2019: 10); that is, the gender “expressed, inferred, and used by a perceiver to classify a referent”.

Our terminology choice is motivated by several observations. First, as noted by Ackerman (2019), the gender expressed by gendered nouns is often defeasible. Further, as Konnelly and Cowper (2020) observe, gender implied by such nouns may be lexically encoded for only some speakers. It should be further pointed out that the relationship between the gendered noun, the referent's gender, and their pronoun of reference is not one to one. Konnelly and Cowper (2020) note that some terms, such as *mother*, may refer not to the birthing parent but to a caretaker, and there are cases where the common noun *dad* is used to refer to an individual whose pronouns of reference are *she/her* (Konnelly and Cowper 2020: notes 33, 34).

translate to processing advantages: bound-variable singular *they* offered a processing advantage over referential singular *they* only with gendered antecedents. Otherwise, bound and referential singular *they* actually both showed processing disadvantages compared to *she/he*.

What is particularly interesting about these results is that the processor is differentially sensitive to gender depending on whether *they* is bound or referential. It has been shown that readers are sensitive to mismatches in the gender of a pronoun and available antecedents, even when there is just one antecedent available (Osterhout and Mobley 1995, Carreiras et al. 1996). When there is a mismatch, it registers as a processing difficulty, either because of a clash in features or because the pronoun is unheralded and the reader is required to accommodate a new referent that may not be readily available. The question is what features *they* has such that it will or will not trigger a mismatch.

We explore how different theories of the featural representation of *they* could account for the processing profile we find. Ultimately, none of these capture the full range of offline and online data we collected. We offer in the final section of this article an alternative formal analysis that distinguishes the way in which gender features are represented on quantified versus referential antecedents. We follow a view suggested in the semantics literature, and closely related to the proposal in Konnelly and Cowper (2020), that quantified antecedents invoke different representations than do referential antecedents. Both antecedents bear formal indices (indicated by numerals such as [1]) which are shared by co-referential and bound pronouns. However, indices are represented and interpreted differently for quantified phrases and referential phrases. In the case of quantified phrases, the index is parsed separately from the quantified phrase as a simple index (Heim and Kratzer 1998) which we will argue optimally bears no gender features, even if the quantified noun phrase itself does. (For convenience we represent gender as the feature [GENDER] below.)

(3) *Quantified antecedent index:*

[Every woman_{GENDER,SG}] [1 [did their₁ homework]]

In contrast, the index on a referential antecedent is parsed as part of the DP, bundled together with any number or gender features associated with the noun.

(4) *Referential antecedent index:*

[The woman]_{1,GENDER,SG} [did their₁ homework]

The idea is that when a pronoun retrieves a referential antecedent, it retrieves all the features in this bundle. We will argue this leads to a clash in a situation like (4) on the view that singular *they* bears (enriched) negative values for gender ([-GENDER]) for some speakers. What is retrieved in bound variables, we suggest, is not the quantified noun phrase itself, but the gender-free index, and bearing no gender feature this does not mismatch with the features of *they*. This proposal has the advantage of holding constant the features of *they*, while at the same time allowing for the fact that *they* is differentially sensitive to gender depending on whether it is bound or referential. While we will ultimately remain uncommitted about how these formal

representations are fully integrated in a theory of sentence processing, we think our contribution highlights the value of combining insights from the formal literature with those of processing theories.

2. BACKGROUND

This section reviews the previous formal linguistic literature and processing studies on singular *they* that serve as background to the work presented here.

2.1 Singular *they* in the formal syntax-semantics literature

The use of singular *they* is undergoing a number of changes in present-day English. What remains invariable, it seems, is that bound-variable uses are readily accepted and have been for centuries. Furthermore, as noted above, there is some evidence that bound-variable singular *they* is possible with antecedents of any gender for even the most conservative speakers. We verify this in the acceptability studies reported below. Referential singular *they*, on the other hand, is subject to a great deal more variation and nuance. Konnelly and Cowper (2020) identify three stages in the expanding use of referential singular *they*. In stage 1, which is that of the most conservative speakers, referential singular *they* is used as an *epicene* pronoun as in (5).

(5) Shhh! The person on the phone with me has lost *their* voice.

In these scenarios the gender of the referent may be unknown or indeterminate (Bodine 1975, Bjorkman 2017) or irrelevant to the communicative goals of the speaker (Moulton et al. 2020). For stage 1 speakers, referential singular *they* is not possible with referents where a form expressing binary gender is both appropriate and known. Such speakers also do not allow singular *they* with antecedent nouns that are gendered (e.g., *sister*). In later stages, speakers use referential singular *they* with referents and antecedents of any gender.³

A critical component of the analyses in Konnelly and Cowper (2020) as well as Bjorkman (2017) is that morphosyntactic features may be contrastive or non-contrastive. If a feature is contrastive in a system, then its absence implies the negation of that feature (or the property that feature represents). For stage 1 speakers, the morpho-syntactic gender features [MASC] and [FEM] are contrastive. Since singular *they* lacks both such features, it implies that gender is unknown or irrelevant. We will describe this as the *epicene* implicature. In later stages, with more innovative speakers, gender features become non-contrastive. The absence of a non-contrastive feature does not trigger the *epicene* implicature and so *they* becomes felicitous in a wider range of discourse contexts with a wider range of antecedent types.

What is of central importance to our studies is the difference between bound and referential singular *they* in the grammar of speakers for whom gender is contrastive. An adequate theory needs to ensure that gender is rendered non-contrastive on bound

³We refer the reader to Konnelly and Cowper (2020) for nuanced differences between Stage 2 and Stage 3 in terms of the feature specification of nouns and pronouns.

variables but remains contrastive on referential singular *they* for conservative speakers (assuming this is the correct description of the facts, which we do verify in Experiments 1a and 2a). Bjorkman (2017) follows Déchaine and Wiltschko (2002) in postulating that bound pronouns may have a smaller structure than referential pronouns. In Bjorkman's (2017) analysis, bound-variable singular *they* can exclude the projection (ϕ P) that houses gender features. We interpret Bjorkman's proposal in the following way: when ϕ P is itself entirely absent, the absence of a specific gender feature does not trigger the epicene implicature. Referential pronouns, in contrast, must include ϕ P; if no specific gender feature is present on ϕ P, then the epicene implicature arises. In a related proposal, Conrod (2019) argues that bound pronouns, unlike referential pronouns, do not involve N movement to D, where gender features are located. Both approaches distinguish bound vs. referential singular *they* in terms of the structure of the pronoun itself.

For Konnelly and Cowper (2020), the bound–referential contrast comes out of differences in the type of antecedent involved. They propose that the entire DP in a quantified antecedent need not inherit the gender features of its common noun restrictor, even if it bears contrastive gender features for Stage 1 English users. The entire DP of a referential antecedent, on the other hand, must bear the gender features of the head noun. Coupled with the additional requirement that “coreference requires that the features of the pronoun match those of its [entire DP] antecedent”, singular *they* will not be possible with a gendered referential antecedent but will be with a gendered quantifier antecedent.

In addition to gender features, the number feature of singular *they* needs to be addressed. From a morpho-syntactic point of view, singular *they* does not bear a singular feature (note that it always triggers subject-verb agreement appropriate for a notional plural: *every/the person said they are/*is here*). One possibility is that *they*, whether interpreted as singular or otherwise, never bears a singular feature (Sauerland et al. 2005). While number is not a dimension along which we manipulated the stimuli of the studies reported below, both for concreteness and to limit the hypothesis space, we follow Sauerland et al. in taking *they* to be inclusive of both plural and singular denotations and in never bearing a morphological or semantic singular feature.

In the next section we turn to psycholinguistic studies which examine the relationship between singular *they* and its antecedent in terms of processing difficulty. Here the precise featural make-up of *they* becomes crucial. We lay out various expectations for processing depending on assumptions about the linguistic representation of *they*. We review the existing evidence in light of these expectations, motivating the experiments to follow.

2.2 Modelling the processing of singular *they*

A number of studies have investigated whether the processor has difficulty when *they* retrieves either a singular antecedent or a gendered antecedent. Underlying these studies are assumptions about what features singular *they* does or does not carry in the first place, such that they would ever cause mismatches. As we saw in the last section, it is not trivial to identify what features *they* carries and this makes it difficult

to make concrete processing predictions. In the following subsections, we outline predictions generated by two different approaches to the gender and number features of *they* and then measure them against the existing processing literature.

As for the crucial distinction between bound and referential *they*, pronouns in English do not carry features that identify them as bound or referential. Nonetheless, since bound-variable singular *they* is the one most widely available across speaker populations, even potentially with gendered antecedents, we might expect processing advantages for such cases. In fact, it has been proposed that for pronouns in general the processor prefers bound interpretations over referential ones (Grodzinsky and Reinhart 1993), although there is no consensus (Frazier and Clifton 2000, Carminati et al. 2002, Koornneef 2008, Koornneef et al. 2011, Cunnings et al. 2014). Koornneef (2008) found that Dutch pronouns were more likely to retrieve quantified over referential antecedents. However, Cunnings et al. (2014) report no such preferences for English, finding instead merely a preference for recency. The question for singular *they* is whether a quantified antecedent offers any advantage in incremental processing, over and above any potential advantages for bound pronouns generally. We discuss an ambiguity theory of bound versus referential *they* below that makes this a viable prediction.

2.2.1 *Underspecification hypothesis*

It has been repeatedly argued that the processing of *they* in all its uses is different from the processing of other pronouns, including singular *he/she/it*. In particular, there is some evidence that *he/she/it* pronouns place a more immediate pressure on the processor to find an antecedent and that they take more resources than *they*. Moxey et al. (2004) found earlier disruptions in the reading of *she/he* lacking a salient singular antecedent compared to *they* lacking a salient plural antecedent. They suggest that the processor does not as immediately need to resolve the antecedent of *they* “possibly because *they* can refer to a wider range of antecedent types than *he/she* can”. Using ERP methods, Filik et al. (2008) found evidence of a cost for unheralded *she/he* but not for *they*. Sanford et al. (2008) found that so-called *institutional they*, which needs no antecedent at all (in referring to implied agents) created no processing costs either. These authors suggest that *they* is an underspecified pronoun, and so will tolerate a wide range of antecedent types. These expectations are often couched in a shallow or good-enough processing model (Ferreira et al. 2002), where *they* would pose no immediate processing difficulty but might require greater resources in later processing (Moxey et al. 2004).

If *they* is indeed underspecified, we may not expect the retrieval of singular or gendered antecedents to pose any processing difficulty, at least in early processing. Moreover, without specifying anything further regarding the difference between bound and referential pronouns (but see below) we do not expect a processing advantage for bound over referential *they*.

2.2.2 *Enriched specification hypothesis*

Another logical possibility is that *they* is specified in some way. We think that the theoretical literature makes some possibly testable predictions in light of the notion of

contrastive features outlined in the last section. The absence of gender and number features on *they* may allow the processor to enrich the features of *they* to include the negative values of these features, that is, [-SG] and [-MASC] and [-FEM]. Such ‘enriched’ features would then clash with a singular antecedent or a gendered antecedent. We intend the enriched specification approach to be one in which the negative values are active as soon as the processor encounters the pronoun. (Rather than, say, one where the enrichment is delayed; in that case, we would have difficulty distinguishing this from the underspecification approach.) As with the underspecification approach, the enriched specification hypothesis is silent on the difference between bound and referential *they*.

2.2.3 Referential vs. bound *they*: ambiguity hypothesis

The above two options concern the role of number and gender. In terms of the bound–referential dimension, we have seen formal theoretical proposals that potentially make interesting processing predictions. As noted above, both Conrod (2019) and Bjorkman (2017) suggest that bound-variable singular *they* has a smaller, simpler structure than referential singular *they*. That means *they* is essentially ambiguous. Upon encountering *they*, readers may access the simplest representation, which is one that is predisposed toward finding a quantified antecedent.⁴ In this case we might expect a processing advantage for bound singular *they* over referential singular *they* whether the antecedent is gendered or not.

While this particular framing of the ambiguity approach predicts an advantage for bound *they*, it does not make predictions about any interactions with gender or number. The two approaches to gender/number outlined above treat bound and referential singular *they* equally. There is a more complex option on the market that we think deserves consideration. One interpretation of the proposal in Bjorkman (2017) is that only referential *they* is enriched with the negative values (since it carries the node that in principle could carry gender features) and hence only it will give rise to feature clashes with a gendered singular antecedent. Bound *they*, on the other hand, would not be enriched, and so would enjoy both an advantage with quantified antecedents and would show no clash with a gendered antecedent. Similar considerations hold for number.

The reports in the formal literature suggest we might find a processing profile consistent with the complex version of the ambiguity hypothesis outlined above: (i) overall that bound singular *they* has a processing advantage over referential singular *they* and (ii) that bound singular *they* has an advantage over referential singular *they* with gendered antecedents.

2.3 Previous studies on processing singular *they*

The experimental record concerning the processing of bound and referential singular *they* is mixed, in part because the extant studies ask rather separate questions and few make explicit assumptions about the featural content of *they*. Doherty and Conklin (2017) investigated the role of the gender stereo-typicality of antecedents, all

⁴We say ‘predisposed’ because referential noun phrases can also antecede bound pronouns, as in sloppy interpretations in ellipsis. We return to this issue in section 6.

referential. Participants in their study showed processing difficulty of *they* with gendered antecedents but no cost for non-gendered antecedents.

Foertsch and Gernsbacher (1997) investigated the impact of both gender and the quantificational status of the antecedent. They measured participants' reading times for passages such as those in (6), where an indefinite antecedent was gender stereotyped (*a truck driver*, as in (6a)), non-gendered (*a runner*, as in (6b)) or was the bare quantifier *anybody*, as in (6c).

(6) *Stimuli from Foertsch and Gernsbacher (1997), Experiment 1*

- a. A truck driver should never drive when sleepy, even if *he/she/they* may be struggling to make a delivery on time. . .
- b. A runner should eat lots of pasta the night before a race, even if *he/she/they* would rather have a steak. . .
- c. Anybody who litters should be fined \$50, even if *he/she/they* cannot see a trashcan nearby. . .

In whole-sentence reading times, sentences containing *they* were read as fast as the sentences containing a pronoun congruent with the gender of the antecedent. With the bare quantifier *anybody* in (6c), *they* actually afforded a reading time advantage over *he* and *she*. In a second experiment, Foertsch and Gernsbacher (1997) tested referential gendered antecedents (*that truck driver*) and found that the sentences with *they* were read more slowly than those with the gender-matching singular pronoun, but with non-gendered referential antecedents (*that runner*) *they* was read as quickly as the singular gendered pronouns. Overall, the results suggest that *they* can resolve to both singular quantified antecedents and non-gendered, singular referential antecedents without apparent difficulty. This outcome is compatible with the underspecification hypothesis for number at least (although they did not compare singular and plural antecedents). Gender, however, appears to cause processing difficulties, but only for referential antecedents. As we saw, neither the underspecification nor the enriched specification hypotheses *alone* predict this interaction. The processing profile that emerges from these studies *is* potentially consonant with the complex ambiguity hypothesis we outlined above: bound singular *they* does not give rise to enriched features specifying negative gender values, unlike referential *they*, and so we do not expect a feature clash in the former.

Caution should be taken in interpreting Foertsch and Gernsbacher's studies. First, they do not directly compare referential vs. bound singular *they* with gendered antecedents in one study. Furthermore, Foertsch and Gernsbacher used a whole-sentence self-paced reading methodology, where each sentence was presented successively in its entirety, making it difficult to locate processing difficulty. One of the key contributions of our studies is to determine whether there is an interaction between gender and quantification in the processing of *they* in a single word-by-word self-paced reading study.

We have further reasons to expect that we might find such an interaction. Ackerman (2018) compared sentences employing *themselves* with gendered and non-gendered antecedents, finding a processing advantage using eye-tracking while reading for both the gendered indefinite (*a mechanic*) and the bare indefinite (*someone*) compared to specific antecedents (i.e., proper names of different gender

bias). Again, the results for gendered indefinites, which can be interpreted as quantificational, fit with the observations in the formal literature that gender interacts with the quantificational vs. referential status of the antecedent, suggesting that bound singular *they* is ‘genderless’ compared to referential singular *they*.

The studies cited above manipulated the gender of the antecedent. Sanford and Filik (2007) investigated the possible clash of number between *they* and a singular antecedent. They suggest that *they* is not initially tolerant of singular antecedents, but the singular antecedent can be subsequently “accommodated in some way” (Sanford and Filik 2007: 172). While tracking participants’ eye-movements, Sanford and Filik presented passages like (7) with singular *someone* or plural *some people* followed by either *them* or a singular *her* downstream.

(7) *Stimuli from Sanford and Filik (2007)*

Mr Jones was looking for the station. He saw [someone/some people] on the other side of the road, so he crossed over and asked [them/her] politely. . .

Their eye-movement data revealed processing difficulties for *they* with a singular antecedent, suggesting that *they* initiates a search for plural antecedents and when the search fails, a cost is incurred. This outcome is compatible with a number of ideas concerning the number features of *they*, including the enriched feature theory elaborated above as long as we allow enriched features to be more defeasible than inherent ones.

One limitation of the experimental studies surveyed above should be emphasized: while suggestive of processing differences between referential and quantified singular *they*, none *systemically* control for the difference between bound-variable and referential singular *they*. All use indefinite antecedents, headed either by an indefinite article (*a* or *some* with a noun phrase complement), or a bare indefinite (*someone* or *anybody*). Indefinites have a notoriously wide range of interpretations, and debate has existed for decades as to whether they are quantificational, referential, or both (Kamp 1981, Heim 1982). A quantificational indefinite is in English usually interpreted existentially and typically requires a licenser, such as negation or a modal. Fodor and Sag (1982) argued that there are also referential uses of indefinites, and this position has reached consensus in the semantics literature, although there are debates about how the referential use arises and is modeled (Reinhart 1997, Winter 1997, Kratzer 1998, Matthewson 1999, Schwarzschild 2002). In simple episodic sentences like (7), used in Sanford and Filik’s (2007) study and Doherty and Conklin’s (2017) study, the indefinite could most naturally be interpreted as referential.⁵ The stimuli in Foertsch and Gernsbacher (1997), on the other hand, are most naturally interpreted with a quantificational interpretation for the indefinite, one in which existential force is interpreted with scope below the deontic modal: (6a) most naturally conveys that it is not compatible with the rules that there exist an *x* such that if *x* is a truck driver, *x* drives when sleepy. Similar remarks apply to the other stimuli in Foertsch and Gernsbacher’s (1997) Experiment 1. This raises the possibility that the differences between the results in Foertsch and Gernsbacher (1997) and

⁵Sanford and Filik (2007) explicitly describe the antecedents in their study as referential.

Sanford and Filik (2007) are due not just to different methodology and antecedent type manipulation, but to differences between the effect that quantificational and referential antecedents may have on processing *they*. Foertsch and Gernsbacher (1997) did not directly compare referential and quantificational antecedents,⁶ nor is it guaranteed that all the indefinites in their stimuli are quantificational, or unambiguously interpreted as such by participants.⁷ Since Sanford and Filik (2007) did not test quantificational antecedents, we do not know whether their finding of a cost for *they*, using finer-grained methodologies than whole sentence reading time, would extend to quantificational and gendered antecedents.

In summary, the processing literature shows that, at least among the English speakers tested, non-gendered antecedents for singular *they* are more acceptable than gendered antecedents (Doherty and Conklin 2017) and that non-gendered antecedents confer upon singular *they* a processing advantage (Foertsch and Gernsbacher 1997, Doherty and Conklin 2017). The suggestive evidence in Foertsch and Gernsbacher (1997) is that these gender and bound-variable properties interact, such that a gendered antecedent has deleterious effects for referential but not for bound-variable singular *they*.

Stepping back, our goal is first, to determine whether this expectation is empirically borne out. Two offline experiments verify the intuitions reported in the syntax-semantics literature about the high acceptability of bound singular *they* with both gendered and non-gendered antecedents, in contrast to referential singular *they*. The self-paced reading experiments (one with non-gendered antecedents, the other with gendered antecedents) then sought to identify whether gender imposes processing difficulty differently for referential vs. bound singular *they*.

2.4 Ensuring bound-variable interpretations

Before turning to the experiments, it is important that we identify how our studies avoid the confounds posed by using an indefinite noun phrase antecedent as was done in the studies documented above. We chose instead to use the universal quantifier *every* in our studies, which is morphosyntactically singular. Universals like *every* are not without complications, since they can indirectly introduce a plural referent – often called the *reference set* or *witness set* (Nouwen 2003, Paterson et al. 2009). *They* can take this plurality as its referent:

- (8) Every person in the room said they were gathered for a nice meal.

They must denote a plurality in (8) since it serves as the argument of the predicate *gather*, which requires a plural subject (*#The person gathered for a nice meal*). Witness-set readings are hard to block. One strategy, following Rullmann (2003), involves contexts that force uniqueness at the level of atoms on the pronoun, as in (9).

- (9) Everyone thinks that they are the smartest person in the world.

⁶Foertsch and Gernsbacher (1997) did perform a between-experiments analysis, showing that gender-matched pronouns had a larger advantage over singular *they* with referential antecedents than quantificational ones.

⁷We do not have access to the full set of stimuli used by Foertsch and Gernsbacher (1997).

If *they* referred to the witness set (the set of people that form the restrictor of the quantifier), then (9) would attribute to each person the belief that all people are the smartest. This is not a felicitous interpretation for (9), and we take it that readers do not pursue such an analysis.

Our experimental stimuli were constructed along these lines in order to force a truly bound singular reading and to block a witness-set reading. For each trial we provide a context sentence that sets up the expectation that the relevant pronoun must refer to a singular atomic individual: in (10), we learn that only one person can win the race. The target sentence (10a), which contains the quantifier and the bound variable, also reinforces the singularity of the pronoun with a singular-enforcing definite description ‘the winner’ in advance of the critical pronoun with which it is identified in a copular relation. The critical pronoun is placed in a post-copula position in a specificational clause. The singular definite description in the pre-copula position (*the winner* in (10a)) is the inverse predicate (Heycock 1992, Moro 1997, den Dikken 2006) and forces the post-copula pronoun to be interpreted as singular.

(10) *Context sentence*: Only one runner could win the race.

a. *Target sentence*: Every runner thought that the winner would be *them/him*.

b. *Target sentence*: The tallest runner hoped that the winner would be *them/him*.

In the experiments we report below, we compare referential DPs (such as *the tallest runner*) as in (10b) to quantificational DPs as in (10a) anteceding *them*; in each case, we use a morphologically singular pronoun (*him*) as the baseline. Note that in the target sentence (10b), the referential antecedent *the tallest runner* contains the superlative adjective *tallest*, which makes it not minimally different from the quantificational antecedent *every runner* in (10a). This modification was necessary in order to facilitate a successful reference. A unique referent of *the runner* without the superlative modifier is not identifiable, as the given context implies that there is more than one runner.⁸

We present results of both acceptability rating and self-paced reading experiments. The first experiment group (Experiment 1a) uses non-gendered antecedents (e.g., *runner*). The second experiment group (Experiment 2a) examines gendered antecedents (e.g., *granddaughter*).

3. EXPERIMENTS 1

Experiment 1a tested the acceptability of singular *they* with non-gendered universally quantified phrases in comparison to non-gendered referential noun phrases.

⁸The presence of the definite noun phrase subject (*the winner*) potentially adds another candidate antecedent for the pronoun, which as a reviewer points out could complicate the processing of these sentences. We acknowledge this limitation. At the same time, this definite noun phrase was present in all conditions. Moreover, in these copular sentences the definite noun phrase subject has the status of a description rather than an individual. Note for instance that the tag is inanimate: “The winner was her, wasn’t it?” (compare to “*The winner was her, wasn’t she?”).

Experiment 1b tested the processing profile of singular *they* with these same two types of antecedents in a self-paced reading (SPR) study. We expected to verify that with truly quantificational antecedents bound singular *they* is more acceptable. If the ambiguity hypothesis holds we expect bound singular *they* to exhibit a processing advantage over referential singular *they*. Furthermore, given that previous literature found that singular *they* with non-gendered antecedents shows improved acceptability and faster processing times compared to gendered antecedents, any degradation in acceptability or difficulty in processing would be most naturally attributable to a sensitivity to number marking, that is, that *they* is less congruent with singular antecedents than *he/she*.

3.1 Experiment 1a

If singular *they* is sensitive to the grammatical number of the antecedent, then we should find sentences containing *them* with singular non-gendered antecedents to be less acceptable than sentences containing *him* with singular non-gendered antecedents. Moreover, if the number on the antecedent has a different effect in the acceptability of singular bound-variable *them* and singular referential *them*, then we should find an interaction between antecedent type (quantificational vs. referential) and pronoun type (*them* vs. *him*).

3.1.1 Materials

Twenty test item sets were constructed as in (11), where a non-gendered universal quantifier subject (QUANT) or a non-gendered referential subject (REF) appeared with a singular gendered pronoun *him* (HE) or *them* (THEY).⁹ These subjects are intended to serve as the antecedent of the pronoun in each target sentence, and were all independently rated as being associated with low gender stereotypicality in Doherty and Conklin (2017). Each item set was thus created crossing two two-level factors, Antecedent (QUANT vs. REF) and Pronoun (THEY vs. HE). Each target sentence was presented with a context sentence as in (11) to further ensure that the relevant pronoun referred to a singular entity.¹⁰

- (11) Only one competitor could win the race.
- | | |
|--------------------------------------------------------------------------------|------------|
| a. <i>Every competitor</i> thought that the winner would be <i>them</i> . | QUANT.THEY |
| b. <i>Every competitor</i> thought that the winner would be <i>him</i> | QUANT.HE |
| c. <i>The youngest competitor</i> hoped that the winner would be <i>them</i> . | REF.THEY |
| d. <i>The youngest competitor</i> hoped that the winner would be <i>him</i> . | REF.HE |

Thirty filler items such as (12) were also included. Each filler item was composed of two sentences: the first sentence contained a gender-stereotyped proper name and an

⁹We chose only *him*, rather than *her*, to instantiate all the singular cases.

¹⁰The full set of materials used in the experiments reported in this article are available on the CJL's website as an appendix, at <http://doi.org/10.1017/cnj.2022.30>.

expression such as *alone* to promote a coreferential interpretation for the subsequent pronoun; the second sentence contained *they* or a singular pronoun that matched the gender of the proper name.

- (12) a. *Bob* was coloring alone in the classroom. While choosing a crayon, *he* refused to pick a bright color.
- b. *Richard* was sleeping alone in the bedroom. After waking up, *they* refused to make some breakfast.

It has been observed in the literature that for some English users (e.g., Stage 1 and 2 in Konnelly and Cowper's 2020 work) *they* cannot generally be used to refer to gendered singular proper nouns. In an experimental setting, Ackerman et al. (2018) found *they* with gender-biased names to be distinctly marked (at least for some participants) when paired with referential *they*. We thus expect that our fillers with *they* will be rated much less acceptable than the ones with singular gendered pronouns.

3.1.2 *Participants and Procedures*

Thirty-six native English users were recruited online using Amazon Mechanical Turk and directed to the experiment on Ibex Farm (Drummond 2013). The age of the participants ranged from 24 to 65, with the mean age at 38. Participants self-reported to be native users of English by answering a survey question at the end of the experiment. Each participant received \$1.50 as compensation for participation upon completion of the experiment.

The test items were distributed over four lists in a Latin-square design so that no participant saw any one item in more than one condition, but all filler items were seen by all participants. Each list contained 20 test items and 30 fillers which were displayed in a randomized order. Participants rated the acceptability of each target sentence from 1 (not acceptable) to 7 (acceptable).

3.1.3 *Results*

The mean ratings and standard errors by condition are provided in Table 1. Also, the distributions of mean ratings of participants and the mean ratings across participants by condition are shown in Figure 1. Each hollow dot represents a mean rating of a participant in a given condition, and each solid dot represents the mean rating across participants in a given condition.

We analyzed the ratings by means of a linear mixed-effects model in R (R Development Core Team, 2020). The lme4 package was used to fit the model (Bates 2005), and the lmerTest package was used to obtain *p*-values (Kuznetsova et al. 2014). In analyses of data obtained from all experiments reported in this paper, we first attempted to fit a maximal random-effects structure with random intercepts and random slopes for participants and items (Barr et al. 2013). If that model did not converge, we fit a model just like the maximal model, but with the random correlation parameter for the interaction term removed for both participants and items. Moreover, the predictors in all

	THEY (<i>them</i>)	HE (<i>him</i>)
QUANT	5.85 (.10)	4.77 (.13)
REF	4.81 (.15)	5.74 (.11)

Table 1: Mean ratings by condition (SE), Experiment 1a

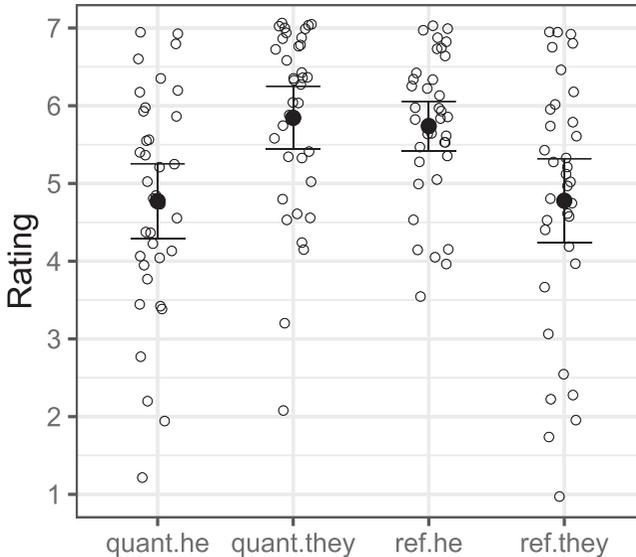


Figure 1: Distributions of mean ratings of participants (hollow dots), and mean ratings across participants with standard errors (solid dots), Experiment 1a

analyses reported here were sum coded, with one of the levels coded as 1, and the other as -1 .¹¹

We fit a mixed model to the ratings with fixed factors of Antecedent (QUANT vs. REF) and Pronoun (THEY vs. HE).¹² We found an interaction between the two factors (Est = 0.51, SE = 0.10, $t = 4.95$, $p < 0.001$). We conducted planned comparisons using pairwise t-tests with Bonferroni adjustment, and compared the ratings on

¹¹A reviewer observes that ideally, an ordinal regression model should be used to analyze the rating data, since technically, Likert scale data are ranked ordinal categories, and not continuous. However, we chose our data analysis method because (i) the use of linear mixed-effects modelling is considered to be the current best practice in the experimental syntax literature for the analysis of numerical judgment data (Schütze and Sprouse 2014); and (ii) there is research that argues that ordinal variables with categories similar to Likert scale can usually be treated as continuous in factor analysis (Robitzsch 2020.)

¹²The formula of the model is: Rating \sim Antecedent*Pronoun + (1+Antecedent*Pronoun|Participant) + (1+Antecedent*Pronoun|Item).

the THEY sentences and the HE sentences in the two Antecedent conditions. According to the planned comparisons, this interaction was due to the fact that the THEY condition had higher ratings than the HE condition in sentences with quantified antecedents (by-participant: $p < 0.01$, by-item: $p < 0.001$), while the reverse was the case in sentences with referential antecedents (by-participant: $p = 0.01$, by-item: $p < 0.001$).

3.1.4 Discussion

Our participants rated *they* sentences with non-gendered quantified antecedent phrases much higher than the ones with non-gendered referential antecedent phrases. This result indicates that our participants accepted *them* as a bound-variable pronoun anteceded by a non-gendered, universally quantified phrase. Participants in fact preferred *them* to the gendered, singular *him* as a bound variable. In contrast, singular gendered pronouns were preferred to *them* as referential pronouns. Note however that referential singular *them* was by no means unacceptable to our participants. Even though the sentences with referential singular *they* were rated lower than the ones with referential singular gendered pronouns, they were rated relatively high (4.80), as high as the sentences with bound singular gendered pronouns (4.77), which is a grammatically possible option.

Comparing the distribution of ratings in the THEY conditions (indicated with hollow dots in Figure 1), while only two participants had mean ratings below 4 in the QUANT.THEY condition, nine participants had mean ratings below 4 in the REF.THEY condition. Thus, more inter-speaker variation is attested in the REF.THEY condition. This finding is consistent with what is reported in Conrod (2019) and Konnelly and Cowper (2020) that speakers range from those who reject the use of referential singular *they* to those who have absolutely no problem with it. Further, upon closer inspection of the data, two participants who had mean ratings below 4 in the REF.THEY condition also had mean ratings below 4 in the QUANT.THEY condition, and seven participants who had mean ratings below 4 in the REF.THEY condition had mean ratings above 4 in the QUANT.THEY condition. Thus, these participants who found singular *they* to be degraded with a referential antecedent found it to be more acceptable with a quantificational antecedent.

The validity of these results is supported by those of the filler sentences. The sentence pairs with *they* were rated much lower (2.95) than the ones with gender-matched singular pronouns (6.42). As with the *they* sentences with referential antecedents, there was variation in the acceptability of *they* sentences with proper-name antecedents among the participants, ranging from those who had very low mean acceptability ratings to those who had very high mean acceptability ratings. But more participants rated the *they* sentences with proper name antecedents below 4 (N=27), in comparison to the *they* sentences with non-gendered referential antecedents (N=9). While some participants who rated the non-gendered referential *they* sentences high also rated the proper name *they* sentences high, many did not. These results confirm the intuition and findings reported in the extant literature that while some users find proper name *they* sentences perfectly acceptable (Conrod 2019, Konnelly and Cowper 2020), for many users, singular *they* anteceded

by gendered proper nouns is less acceptable than with referential DPs (Bjorkman 2017, Ackerman et al. 2018).

3.2 Experiment 1b

Since Experiment 1a showed that with non-gendered antecedents, referential and bound singular *they* were acceptable at different rates, we asked whether this difference appeared in online processing. The expectation is that referential singular *they* will show elevated reading times compared to bound-variable singular *they* (when measured against the baseline *him*). We should thus find an interaction between antecedent type (quantificational vs. referential) and pronoun type (*them* vs. *him*).

3.2.1 Materials

The materials were similar to the ones used in Experiment 1a, crossing two two-level factors, Antecedent (QUANT vs. REF) and Pronoun (THEY vs. HE), yielding four experimental conditions. The test sentences in Experiment 1b, however, were made to be longer so that the sentences do not end with the target region containing the critical pronoun. Also, the definite description in the embedded specificational clause began with *the one who* to ensure a singular interpretation of the post-copula pronoun. Excluding the context sentences, the target sentences were divided into ten regions, with region 1 containing the antecedent and region 7, the target region, containing the critical pronoun, as illustrated (13).¹³

(13) Only one competitor could win the race.

- a. /₁ Every competitor /₂ thought that /₃ the one who /₄ would win /₅ the race /₆ would be /₇ them when /₈ the times /₉ were finally /₁₀ announced. QUANT.THEY
- b. /₁ Every competitor /₂ thought that /₃ the one who /₄ would win /₅ the race /₆ would be /₇ him when /₈ the times /₉ were finally /₁₀ announced. QUANT.HE
- c. /₁ The youngest competitor /₂ thought that /₃ the one who /₄ would win /₅ the race /₆ would be /₇ them when /₈ the times /₉ were finally /₁₀ announced. REF.THEY
- d. /₁ The youngest competitor /₂ thought that /₃ the one who /₄ would win /₅ the race /₆ would be /_{6'} him when /₈ the times /₉ were finally /₁₀ announced. REF.HE

3.2.2 Participants and Procedures

194 native English users, who did not participate in Experiment 1a, were recruited online using Amazon Mechanical Turk and directed to the experiment on Ibx Farm. The age of the participants ranged from 20 to 74, with the mean age at 40.

¹³The target region contains the critical pronoun and *when* (*them/him when*). The critical pronoun marks the end of a clause. Thus, if the target region only contained the critical pronoun, the reader might treat this region as the end of the sentence, possibly resulting in longer reading time due to wrap-up, especially in early trials where punctuation conventions in the study are not clear. With *when* in the target region, which signals the beginning of a subsequent clause, the reader is likely to be aware that the sentence does not end with the critical pronoun.

Participants self-reported to be native users of English by answering a survey question at the end of the experiment. Each participant received \$1.50 as compensation upon completion of the experiment.

Twenty item sets like (13) were distributed over four lists in a Latin-square design. In addition, each list contained a set of 40 fillers. The sentences were presented on Ixus Farm in a uniquely generated random order for each participant, using the moving-window paradigm (Just et al. 1982). After reading the context sentence, participants advanced to the next region by pressing on the space bar. No region could be displayed more than once. After each experimental sentence was read, a comprehension question was presented, which could be answered by pressing one key for 'yes' or another key for 'no'. The comprehension questions tested participants' understanding of the sentence, but not their interpretation of the critical pronoun. The comprehension question for the item in (13) is in (14).

(14) Was there going to be a winner of a race?

3.2.3 Results

Participants with low comprehension question response score (<50%) and extremely fast reading speed per region (<50ms) were excluded. This resulted in eliminating one participant from analysis due to a low comprehension question response score (36%), leaving 193 participants. Further, using the trimr package (Grange 2015), reading times of a region that were 10 standard deviations above the mean were removed, in order to exclude extreme outliers from analysis. Altogether, this resulted in removing 0.5% of the observations from the data.

The grand mean comprehension question response score on test sentences was 89%. The mean proportions of correct responses for the comprehension questions are reported in Table 2. The comprehension questions tested participants' attention to the overall sentence content, and the results show no impact of the manipulated factors on comprehension generally.

Mean raw reading times and mean residual reading times (RRTs) by condition for the regions of analysis are reported in Table 3. These represent reading times for all data, regardless of whether the comprehension question was answered correctly. The regions of analysis are Region 7 (the target region), Region 8 (the spillover region), and Region 9. We calculated RRTs using character length from the entire dataset (including fillers) to estimate the reading time for each region for each participant (Ferreira and Clifton 1986, Trueswell and Tanenhaus 1994, Phillips 2006). The graph in Figure 2 summarizes mean RRTs by condition for the regions of analysis.

We analyzed each region's RRTs with a mixed model, with a random-effects structure as described for Experiment 1a.¹⁴

In analyzing the RRTs of region 7 (target region), we found no main effect or interaction. In region 8 (spillover region), the analysis showed a main effect of

¹⁴The formula of the model applied to each region of analysis is: $RRT \sim \text{Antecedent} * \text{Pronoun} + (1 + \text{Antecedent} * \text{Pronoun} | \text{Participant}) + (1 + \text{Antecedent} * \text{Pronoun} | \text{Item})$.

	THEY (<i>them</i>)	HE (<i>him</i>)
QUANT	.89 (.003)	.88 (.003)
REF	.88 (.003)	.90 (.003)

Table 2: Proportion of Correct Responses (SE), Experiment 1b

Region		RT			RRT		
		7	8	9	7	8	9
QUANT	THEY	371 (8)	401 (8)	397 (7)	-36 (6)	-41 (6)	-44 (6)
	HE	365 (8)	394 (8)	401 (7)	-27 (6)	-49 (7)	-38 (6)
REF	THEY	382 (8)	416 (8)	402 (7)	-24 (7)	-26 (6)	-36 (5)
	HE	364 (8)	392 (7)	393 (7)	-27 (7)	-50 (6)	-47 (6)

Table 3: Mean Raw Reading Times (SE) and Mean RRTs (SE) in ms, Experiment 1b

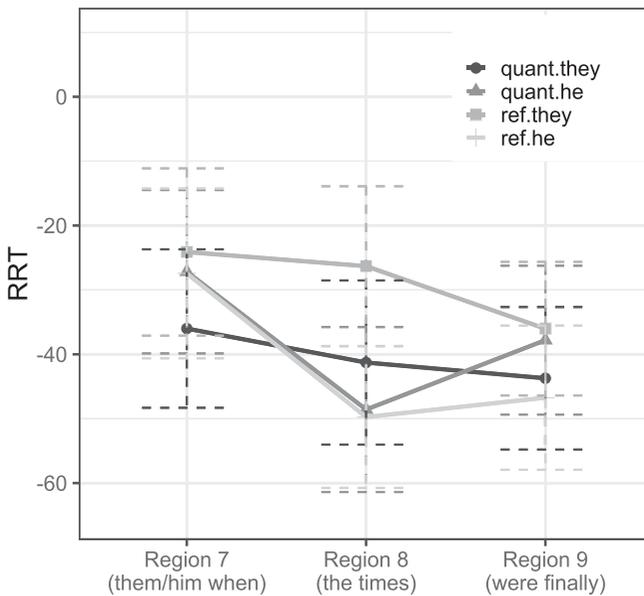


Figure 2. Mean RRTs and standard errors for the regions of analysis, Experiment 1b

Pronoun (Est = 7.62, SE = 3.25, $t = 2.34$, $p < 0.05$), such that overall the THEY condition showed slower reading times than the HE condition. In region 9, the analysis did not reveal any effect.

3.2.4 Discussion

Bound-variable and referential singular *they* did not differ in reading time measures, contrary to the differences found in acceptability. Moreover, bound singular *they*, like referential *they*, incurs a processing cost in the spillover region, revealed by the main effect of Pronoun that persisted across REF and QUANT conditions. This finding is important for several reasons. First, recall that Foertsch and Gernsbacher (1997) did *not* find slower reading times for sentences containing singular *they* and non-gendered referential antecedents. Their finding was called into question on methodological grounds by Sanford and Filik (2007), who showed that a finer-grained measure of processing difficulty – eye-tracking – *does* reveal a processing cost for singular *they* with non-gendered referential antecedents. As pointed out by a reviewer, the two studies are asking different questions. Foertsch and Gernsbacher (1997) are investigating gender processing, while Sanford and Filik (2007) are investigating number processing. The main conclusions of the two studies are therefore not necessarily mutually incompatible. Nonetheless, the results of our Experiment 1b show that self-paced reading is, like eye-tracking, sensitive enough to detect a processing cost. Second, these results demonstrate that even though bound singular *they* is preferred in offline judgments to *he/him* with non-gendered antecedents, it nonetheless poses a processing cost, one that appears to be overcome in reflective judgments without entailing reduced acceptability. But bound *they* and referential *they* have the same processing profile here. This does not bear out predictions we derived from the ambiguity hypothesis; namely, that the linguistically simpler bound *they* would be accessed first and pose no processing problems upon retrieving a quantified antecedent, while retrieving a referential antecedent might require re-analyzing the pronoun as referential. We return to the significance of the absence of such a finding in section 5.

4. EXPERIMENTS 2

Experiment 2a investigated the acceptability of singular *they* with gendered quantificational and referential antecedents. As noted in footnote 2, by *gendered* we mean both nouns like *grandson* and *woman* and gender-stereotyped nouns like *nurse* and *surgeon*. As with Experiment 1a, Experiment 2a was an acceptability rating study, to confirm intuitions reported in the literature that gendered antecedents do not reduce the acceptability of bound singular *they* to the same extent, if at all, as they do for referential singular *they*. Experiment 2b was designed to compare the processing profile of singular *they* with quantified and referential gendered antecedents. The previous literature found that gendered antecedents generally reduce the processing ease of referential singular *they* (Foertsch and Gernsbacher 1997, Doherty and Conklin 2017, Ackerman 2018, Ackerman et al. 2018). Building on Foertsch and Gernsbacher (1997), however, we expect that bound-variable singular *they* will not show this same sensitivity to gender, and will be read with less difficulty than referential singular *they*.

4.1 Experiment 2a

If the gender of the antecedent plays a role in the acceptability of singular *they*, then we should find sentences containing *them* with singular gendered antecedents to be less acceptable than sentences containing *him/her* with the same type of gendered antecedents. Moreover, if a gendered antecedent has a different effect in the acceptability of singular bound-variable *them* and singular referential *them*, then we should find an interaction between antecedent type (quantificational vs. referential) and pronoun type (*them* vs. *him/her*).

4.1.1 Materials

Twenty test items similar in form to the materials used in Experiment 1a were constructed as in (15). Like the items in Experiment 1a, the antecedent phrases were either quantified (QUANT) or referential (REF). In all cases, the nouns were gendered.¹⁵ Half of the stimuli use antecedents associated with female gendered individuals and the other half with male gendered individuals. The pronoun was either *them* or whichever singular pronoun (*him/her*) was appropriate to the gender of the antecedent.

- (15) Only one policeman could win the race.
- | | | |
|----|----------------------------------------------------------------------------|------------|
| a. | <i>Every policeman</i> thought that the winner would be <i>them</i> . | QUANT.THEY |
| b. | <i>Every policeman</i> thought that the winner would be <i>him</i> . | QUANT.S/HE |
| c. | <i>The youngest policeman</i> hoped that the winner would be <i>them</i> . | REF.THEY |
| d. | <i>The youngest policeman</i> hoped that the winner would be <i>him</i> . | REF.S/HE |

In addition, 30 filler items were included which were used in Experiment 1a.

4.1.2 Participants and Procedures

Thirty-seven native English users, who did not participate in Experiments 1a or 1b, completed the experiment online, receiving \$1.50 for compensation. They were recruited using Amazon Mechanical Turk and redirected to the experiment on Ibex Farm. The age of the participants ranged from 23 to 59, with the mean age at 37. Participants self-reported to be native users of English by answering a survey question at the end of the experiment.

Twenty item sets as in (15) were distributed over four lists in a Latin-square design. In addition, each list contained the same set of 30 fillers.

4.1.3 Results

The mean ratings and standard errors by condition are provided in Table 4. The distributions of mean ratings of participants and the mean ratings across participants by condition are shown in Figure 3. Each hollow dot represents a mean rating of a

¹⁵16 belong to the so-called definitionally gendered class (e.g., *policeman*, *actress*, *granddaughter*) while four were gender-stereotyped (e.g., *nurse*, *surgeon*, *secretary*).

	THEY (<i>them</i>)	HE (<i>him</i>)
QUANT	5.46 (.11)	5.80 (.11)
REF	4.21 (.15)	6.05 (.10)

Table 4: Mean ratings by condition (SE), Experiment 1b

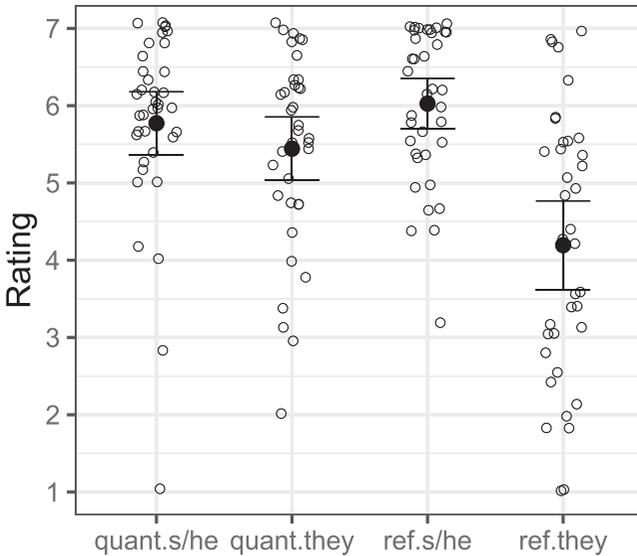


Figure 3. Distributions of mean ratings of participants (hollow dots), and mean ratings across participants with standard errors (solid dots), Experiment 2a

participant in a given condition, and each solid dot represents the mean rating across participants in a given condition.

We analyzed the mean ratings with a mixed model, with a random-effects structure as described for Experiment 1a.¹⁶ We found a main effect of Pronoun (Est = -0.54, SE = 0.10, $t = -5.37$, $p < 0.001$) and a main effect of Antecedent (Est = 0.23, SE = 0.06, $t = 3.92$, $p < 0.001$), such that overall the sentences with *him/her* (mean rating: 5.93) were rated higher than the ones with *them* (mean rating: 4.84), and the sentences with quantified antecedent phrases (mean rating: 5.63) were rated higher than the ones with referential antecedent phrases (mean rating: 5.13). Crucially, we found an interaction between the two factors (Est = 0.37, SE = 0.05, $t = 7.70$, $p < 0.001$). Planned comparisons using pairwise t-tests

¹⁶The formula of the model is: Rating ~ Antecedent*Pronoun + (1+Antecedent+Pronoun|Participant) + (1+Antecedent+Pronoun|Item). As the full model did not converge, the random correlation parameter for the interaction term was removed for both participants and items.

with Bonferroni adjustment revealed that this interaction was due to the fact that for sentences in the REF condition, the *them* sentences were rated lower than the ones with singular gendered pronoun (by-participant: $p < 0.001$, by-item: $p < 0.001$), while in the QUANT condition, sentences with *them* were rated as high as the ones with a singular gendered pronoun (by-participant: $p = 1.00$, by-item: $p = 0.79$).

4.1.4 Discussion

The results of Experiment 2a are similar to those of Experiment 1a. The participants rated *them* sentences with gendered quantified antecedent phrases much higher than the ones with gendered referential antecedent phrases. Participants also rated sentences with singular gendered pronouns much higher than the ones with *them* in the REF condition. One notable difference, however, is that with gendered quantified antecedents, sentences with singular gendered pronouns were rated just as high as the ones with *them*, whereas they were rated lower than the sentences with *them* in Experiment 1a. Another difference is that the *them* sentences with referential antecedents in Experiment 2a (4.21) are numerically rated lower than the ones in Experiment 1a (4.81). We interpret this as a cumulative effect of number and gender on the acceptability of referential singular *them*: neither the number nor the gender of the antecedent is expressed by the pronoun. These results taken together suggest that gender plays a role only in the acceptability of referential singular *they* as expected, but it plays a different role for bound-variable singular *they*.

Comparing the distribution of ratings in the THEY conditions (indicated with hollow dots in Figure 3), five participants had mean ratings below 4 in the QUANT.THEY condition, and 17 participants had mean ratings below 4 in the REF.THEY condition. Thus, as in Experiment 1a, more inter-speaker variation is attested in the REF.THEY condition than in the QUANT.THEY condition in Experiment 2a, with speakers ranging from those who reject the use of referential singular *they* to those who accept it (Conrod 2019, Konnelly and Cowper 2020). Further, upon closer inspection of the data, while the same five participants had mean ratings below 4 in both the REF.THEY and the QUANT.THEY condition, 13 participants who had mean ratings below 4 in the REF.THEY condition had mean ratings above 4 in the QUANT.THEY condition. Thus, as with the results in Experiment 1a, many participants in Experiment 2a who found singular *they* to be degraded with a referential antecedent found it to be more acceptable with a quantificational antecedent.

The results for filler items were similar to Experiment 1a. The sentence pairs with *they* were rated much lower (3.69) than the ones with gender-matched singular pronouns (6.39). Looking at the *they* sentences more closely, there was a variation in the acceptability among the participants, with 15 participants having mean ratings above 4, and 22 participants below 4. As in Experiment 1a, while some participants who rated the gendered referential *they* sentences high also rated the proper-name *they* sentences high, many participants rated the proper-name *they* sentences lower than the gendered referential *they* sentences, resulting in lower mean rating for the proper-name *they* sentences (3.69) than the gendered referential *they* sentences

(4.21). As in Experiment 1a, the filler results in Experiment 2a confirm the intuition and findings reported in the extant literature that for many speakers, singular *they* anteceded by gendered proper nouns is less acceptable than referential DPs (Bjorkman 2017, Ackerman et al. 2018), and at the same time, there are speakers who find no problem at all with proper-name *they* sentences (Conrod 2019, Konnelly and Cowper 2020).

4.2 Experiment 2b

While the offline acceptability of bound singular *they* with a gendered quantifier antecedent was high, particularly compared to referential *they*, the question arises whether this leads to any processing differences. If the gender of the antecedent plays a role in the processing of singular *they*, then we should find *they/them* with singular gendered antecedents to be more difficult to process (increase in reading time) than *him/her* with singular gendered antecedents. Moreover, if the gender of the antecedent has different effects on the processing of singular bound-variable *them* versus singular referential *them*, then we should find an interaction between antecedent type (quantificational vs. referential) and pronoun type (*them* vs. *him/her*).

4.2.1 Materials, Participants and Procedures

The materials were similar to the ones used in Experiment 1b, except that the antecedent noun phrases were gendered, as in (16). Just as in Experiment 1b, each item set represented four conditions, crossing two two-level factors, Antecedent (QUANT vs. REF) and Pronoun (THEY vs. S/HE).

(16) Only one policeman could win the race.

- a. $/_1$ *Every policeman* $/_2$ thought that $/_3$ the one who $/_4$ would win $/_5$ the race $/_6$ would be $/_7$ *them* when $/_8$ the times $/_9$ were finally $/_{10}$ announced. QUANT.THEY
- b. $/_1$ *Every policeman* $/_2$ thought that $/_3$ the one who $/_4$ would win $/_5$ the race $/_6$ would be $/_7$ *him* when $/_8$ the times $/_9$ were finally $/_{10}$ announced. QUANT.S/HE
- c. $/_1$ *The youngest policeman* $/_2$ thought that $/_3$ the one who $/_4$ would win $/_5$ the race $/_6$ would be $/_7$ *them* when $/_8$ the times $/_9$ were finally $/_{10}$ announced. REF.THEY
- d. $/_1$ *The youngest policeman* $/_2$ thought that $/_3$ the one who $/_4$ would win $/_5$ the race $/_6$ would be $/_7$ *him* when $/_8$ the times $/_9$ were finally $/_{10}$ announced. REF.S/HE

Twenty item sets like (16) were created and distributed over four lists in a Latin-square design. In addition, each list contained a set of 40 fillers that were used in Experiment 1b. The sentences were presented in Ibx Farm, following the same procedure as Experiment 1b. 168 native English users, who did not participate in Experiments 1a, 1b, or 2a, were recruited online using Amazon Mechanical Turk and directed to the experiment on Ibx Farm. The age of the participants ranged from 25 to 72, with the mean age at 43. Each participant received \$1.50 as compensation upon completion of the experiment.

4.2.2 Results

Just as in Experiment 1b, participants with low comprehension question response score (<50%) and extremely fast reading speed per region (<50ms) were excluded. This resulted in eliminating two participants: one was due to low comprehension question response score (47%) and another was due to extremely fast average reading speed per region (47ms). This left 166 participants for analysis. Reading times that were 10 standard deviations above the mean were also removed. This resulted in removing 1.2% of the observations from the data for analysis.

The grand mean comprehension-question response score on test sentences was 91%. The mean proportions of correct responses for the comprehension questions of the test items are given in Table 5.

Mean raw reading times and mean RRTs by condition for the regions of analysis are reported in Table 6. These represent reading times for all data, whether the comprehension question was answered correctly or not.

The graph in Figure 4 summarizes mean RRTs by condition for the regions of analysis for all data.

As in the analysis performed in Experiment 1b, here we analyzed each region's RRTs with a mixed model, with a random-effects structure as described for Experiment 1a.¹⁷

In region 7 (target region), the analysis revealed a main effect of Antecedent (Est = -14.70, SE = 4.41, $t = -3.34$, $p < 0.01$) such that overall the sentences with quantificational antecedents (mean RRT: -34 ms, mean raw RT: 369 ms) had faster reading times than the ones with referential antecedents (mean RRT: -5 ms, mean raw RT: 398 ms). The analysis also revealed an interaction between Antecedent and Pronoun (Est = -9.06, SE = 4.46, $t = -2.03$, $p < 0.05$). According to the results of planned comparisons using pairwise t-tests with Bonferroni adjustment, the interaction is due to the THEY condition having longer reading time than the S/HE condition with a referential antecedent (by-participant: $p < 0.05$, by-item: $p = 0.05$). In contrast, the two pronoun conditions showed similar reading times with a quantificational antecedent (by-participant: $p = 1.00$, by-item: $p = 1.00$).

In the analysis of RRTs in region 8 (spillover region), we found a main effect of Antecedent (Est = -9.66, SE = 3.65, $t = -2.65$, $p < 0.01$) and a main effect of Pronoun (Est = 20.39, SE = 4.39, $t = 4.65$, $p < 0.001$). Overall, the referential condition (mean RRT: -18 ms, mean raw RT: 418 ms) had a longer reading time than the quantificational condition (mean RRT: -39, mean raw RT: 399 ms), and the THEY condition (mean RRT: -8 ms, mean raw RT: 429 ms) had a longer reading time than the S/HE condition (mean RRT: -49, mean raw RT: 388 ms). The analysis in region 9 did not reveal any effect.

4.2.3 Discussion

The results of Experiment 2b reveal a difference between quantificational and referential antecedents: with gendered antecedents referential singular *they* exhibits a

¹⁷The formula of the model applied to each region of analysis is: $RRT \sim \text{Antecedent} * \text{Pronoun} + (1 + \text{Antecedent} * \text{Pronoun} | \text{Participant}) + (1 + \text{Antecedent} * \text{Pronoun} | \text{Item})$.

	THEY (<i>them</i>)	S/HE (<i>him/her</i>)
QUANT	.93 (.003)	.89 (.003)
REF	.93 (.003)	.91 (.003)

Table 5: Proportion of Correct Responses (SE), Experiment 2b

Region		RT			RRT		
		7	8	9	7	8	9
QUANT	THEY	372 (8)	414 (9)	401 (11)	-37 (6)	-24 (7)	-36 (10)
	S/HE	365 (8)	384 (8)	389 (7)	-30 (6)	-54 (7)	-45 (5)
REF	THEY	419 (13)	444 (11)	411 (8)	9 (11)	8 (9)	-25 (6)
	S/HE	377 (8)	392 (8)	396 (9)	-19 (5)	-43 (5)	-41 (7)

Table 6: Mean Raw Reading Times (SE) and Mean RRTs (SE) in ms, Experiment 2b

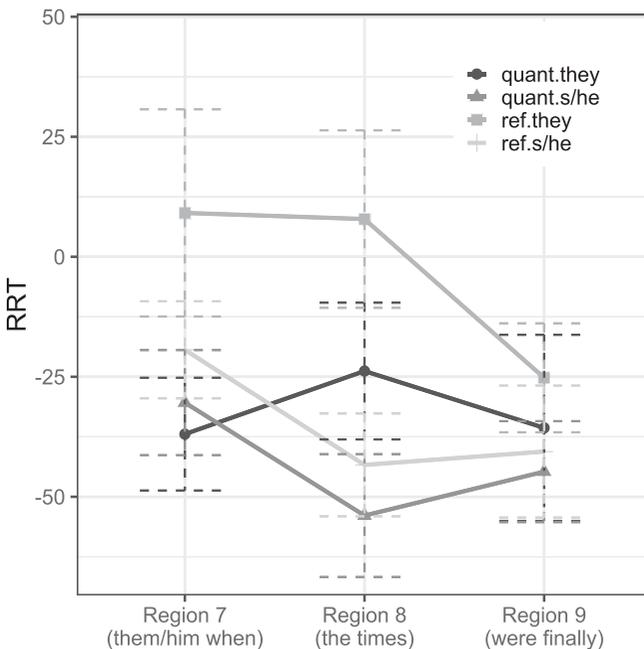


Figure 4. Mean RRTs and standard errors for the regions of analysis, Experiment 2b

processing difficulty in comparison to *him/her*, while bound singular *they* is processed just as easily as the singular gendered pronoun in the target region. This is the gender-quantifier interaction, presaged in the studies by Foertsch and Gernsbacher (1997) and Ackerman (2018). In the general discussion we turn to how to account for the differential sensitivity to gender by bound versus referential *they*, right at the point of encountering the pronoun.

In addition to the differences between bound and referential *they*, there was still a residual processing cost for both bound and referential singular *they* compared to the singular gendered pronouns. At the spillover region there was a main effect that penalized *they* across the board. Note that the same main effect of Pronoun was found in the spillover region of Experiment 1b with non-gendered antecedents. What we are seeing, then, is that even the highly acceptable bound singular *they* can exhibit a small processing cost in comparison to a singular gendered pronoun. It is possible that this is due to a consistent, if weak and temporary, cost for singular antecedents for *they/them*. This is consistent with a theory in which *they* cues a search for a non-singular antecedent, as on the feature enrichment hypothesis; the retrieved antecedent mismatches in number, thus registering as a slowdown.

5. GENERAL DISCUSSION OF EXPERIMENTS

In terms of processing difficulty, we found differences between quantificational and referential singular *they* only with gendered antecedents. Otherwise, singular *they* exhibited a slowdown in the spillover region with both quantified and referential antecedents, both gendered and non-gendered. The latter finding is not expected on the underspecification hypothesis we laid out in section 2.2, which suggests *they* would readily tolerate singular antecedents. Instead, it is in line with Sanford and Filik (2007), who suggest that *they* launches a search for a non-singular antecedent. This is what the enrichment hypothesis predicts when applied to both bound and referential *they*. The across-the-board spillover effect is not compatible with the ambiguity hypothesis. The way we spelled out that hypothesis predicted that bound-variable *they* is both more readily accessed and does not trigger enrichment of number and gender features. We would not have expected bound singular *they* to pose any processing difficulties, especially with non-gendered antecedents, contrary to fact.

Where we found a bound–referential difference in processing was in interaction with gender. This suggests that there is indeed a processing advantage for bound over referential singular *they* but it cannot be a wholesale advantage. This is consonant with the expectations of the ambiguity hypothesis only as long as number and gender are distinguished. This would require number to be enriched on bound-variable *they* (causing a number clash) but not gender (avoiding a gender clash). In principle such representations could be constructed, given a highly articulated syntax with separate projections housing number and gender. Nonetheless, we think divergences between the offline and online results speak against such a move. This move would require that feature enrichment be defeasible so that while a [-SG] enriched

feature triggers a clash, in reflective judgment it could be cancelled on bound *they*, but not on referential *they*, and allow only bound *they* to be highly acceptable with a singular antecedent. Whether that is itself a plausible process, we do not know, but it leaves unanswered why [-SG] would be defeasible only on bound *they*.

Even if the ambiguity hypothesis can be re-engineered to account for the differential interaction of number and gender with referential versus quantified antecedents, it suffers from a more general failure in light of the processing results of Experiment 1b. The ambiguity hypothesis more generally hinges on the assumption that readers will pursue a bound interpretation before a referential one, and upon retrieving a referential antecedent, would require re-analysis. There is no hint of this in Experiment 1b: that is, we found no additional cost for referential *they* compared to bound-variable *they*. So while the ambiguity account is successful in offering a place to locate the bound–referential distinction, it is not successful in accounting for either the offline or online results. To summarize with respect to the processing hypotheses: neither the underspecification hypothesis nor the ambiguity hypothesis were borne out; the enriched specification hypothesis was borne out for number. In section 6, we offer an alternative formal representation for bound vs. referential singular *they* that can capture the interaction between gender and quantification.

An interesting outcome of the studies was that the offline results were in alignment with the online results in some cases but not others. The online results for referential singular *they* were directly reflected by the offline results. In the acceptability judgment task, we found that referential singular *they* is less acceptable than referential singular gendered pronouns with both gendered and non-gendered antecedents. In both cases *they* was processed more slowly than *he/she*. The offline and online results for bound singular *they* were not in such neat alignment. For bound singular *they*, we found a processing delay with both gendered and non-gendered antecedents in the spillover region, just as with referential singular *they*. However, unlike referential singular *they*, we did not find any processing delay with either gendered or non-gendered antecedents for bound singular *they* at the pronoun region. In the offline acceptability judgment task, we found that bound singular *they* is just as acceptable as bound singular gendered pronouns with gendered antecedents. With non-gendered antecedents, it was even more acceptable than singular gendered pronouns. These findings suggest that while the antecedents' number incurs processing cost for bound singular *they*, just as for referential singular *they*, the gender of the antecedent does not. Nonetheless, it appears that this difficulty incurred by the antecedents' number is quickly overcome, as reflected by the offline results.

This kind of mismatch between online processing cost and offline acceptability can be found elsewhere in the literature: it has been shown that while singular gendered pronouns (*he* or *she*) that mismatch in gender with gender stereotyped antecedents (*the nurse* or *the surgeon*) incur processing difficulty, they do not result in degraded acceptability (Kreiner et al. 2008). On the other hand, singular gendered pronouns that mismatch in gender with gendered antecedents (*the policeman*, *the granddaughter*) not only incur processing difficulty but also degrade acceptability. These findings suggest that the gender evoked by stereotype may be temporary, only affecting online processing, but the gender evoked by lexical properties of the

antecedent persists, affecting both the online processing and the offline acceptability judgments. In a similar vein, the processing and acceptability mismatch of singular *they* that we found can be taken to mean that number is temporarily evoked in processing singular *they* as a bound variable, but the initial processing difficulty is overcome in reflective judgments.

In summary, results of the four studies confirm that bound-variable singular *they* enjoys an advantage over referential singular *they*, but not a wholesale one. Rather, the picture is nuanced. In processing, bound and referential singular *they* both show disadvantages compared to singular pronouns – which we took to be a type of number clash. Only with gendered antecedents did the bound-variable singular *they* offer a processing advantage, suggesting an interaction between gender and quantification. We detailed how the ambiguity hypothesis, which predicted advantages for bound over referential singular *they*, was not fully successful in accounting for the full pattern of outcomes. We end this article with a theoretical re-appraisal of the formal representation of bound vs. referential *they*, and test it against the offline and online results.

6. A THEORETICAL RE-APPRAISAL

The ambiguity hypothesis placed the distinction between bound and referential singular *they* on the pronoun. An alternative, already suggested by Konnelly and Cowper (2020), is that the locus of that difference is in the antecedent itself. We would like to put a version of this approach on the table. This version deploys representations that involve binder indices at the syntax-semantic interface in the style of Kratzer (2009) and Sudo (2012, 2014).

In the widely-adopted formal implementations of Heim and Kratzer (1998), quantifier phrases (QPs) bear a numerical index n . To bind a co-indexed pronoun they undergo quantifier raising (QR) leaving a co-indexed trace in their base position. In this framework, the index on the QP is then re-parsed as a separate projection, as shown in (17):

- (17) Every student₈ [t₈ did her₈ homework]
 \Rightarrow Every student [8 [t₈ did her₈ homework]]

There are refinements explored in Sudo (2012) and Kratzer (2009) in which the binder carries interpreted ϕ -features (which includes gender and number), as in (18). Like other ϕ -features, these are interpreted as presuppositions. They impose restrictions on values of the bound pronoun so that it bears the same features and is spelled out accordingly, that is, as *her* in (18).¹⁸

- (18) Every student $8_{[FEM,SG]}$ [t₈ did her₈ homework]

Locating gender features on a binder index affords us a way to regulate the presence of gender features on bound pronouns differently from referential pronouns (we return below to the issue of number). On this approach, singular *they* can only

¹⁸We remain agnostic as to whether there is a morphological process of feature transmission (Kratzer 2009).

arise for the most conservative/Stage 1 users if the binder index does not bear a gender feature. In order to allow for this even when the restrictor noun in the QP is gendered, we propose that Binder Indices are subject to the economy constraint in (19).

(19) BINDER INDEX ECONOMY

Keep ϕ -features on binder indices to a minimum.

Assuming that *they* is unspecified for number and gender features, BINDER INDEX ECONOMY will favor representations like (20):

(20) Every student/girl t_8 [t_8 ... they $_8$...]

There are two sources of independent justification for BINDER INDEX ECONOMY. The first is that across languages, bound pronouns often show fewer morpho-syntactic distinctions than free pronouns. Syntactically bound elements (like reflexive anaphora) often exhibit fewer contrasts in number, person and gender dimensions than non-bound pronouns (e.g., the simple anaphors like *zich* and *sig* in Dutch and Icelandic, and long-distance anaphors such as *ziji* and *caki* in Chinese and Korean). A natural way to capture this state of affairs is to pressure binder indices to make fewer featural contrasts than the pronoun system itself.

A second motivation for BINDER INDEX ECONOMY relates to the role of ϕ -features in a language's agreement system. There are views in which binder indices are housed on functional heads that might possibly be agreeing verbal functional projections (Adger and Ramchand 2005, Kratzer 2009). Kratzer argues that the ϕ -features on binder indices are intimately connected to the way the language's morphological agreement system operates. If so, then binder indices should look more like English verbal agreement than like pronouns. English verbal agreement morphology is not rich, and does not expone gender. It is a small step to imagine that encoding gender features in the syntactic representation of functional nodes is thus militated against, and this is precisely the intention of BINDER INDEX ECONOMY. Certainly binder indices *can* include gender features (this is required for a sentence like *Every girl did her homework*); but the lack of gender contrasts in the agreement system generally might make gender more susceptible to BINDER INDEX ECONOMY than other features. This makes a useful cross-linguistic prediction. Many languages *do* show gender distinctions in some or all of their verbal morphology; we do not expect these languages to easily tolerate minimal binder indices, and so we do not expect them to have an equivalent of bound singular *they*. Thorough cross-linguistic work is needed in this area to test this prediction.

In contrast to quantificational antecedents, a referential DP need not undergo QR in order to be co-valued with a pronoun. A referential phrase may merely bear the same index as a downstream pronoun and be co-referential. It does not need to generate a binder index, and without a binder index, BINDER INDEX ECONOMY would not apply.¹⁹ In examples such as (21), therefore, *they* will not match the gender

¹⁹Referential DPs *can* introduce a binder index, for instance, for sloppy readings in ellipsis. If referential antecedents generate a binder index, and if that index is subject to BINDER INDEX ECONOMY, we might ask whether that option would improve the acceptability and processing of

expectations set up by the gendered antecedent (*the policeman*), for speakers who disprefer referential singular *they*.

(21) The policeman_{8[MASC,SG]} [... !they₈ ...]

Whether or not a speaker finds referential singular *they* degraded, the proposed BINDER INDEX ECONOMY should be active for all speakers in constraining variable-binding representation. Hence, while there may be some variation regarding the acceptability of referential singular *they*, we predict less variation for bound singular *they*. The results in Experiments 1a and 2a are compatible with this prediction: many participants who rated referential singular *they* low assigned higher ratings to bound singular *they*.

We think representations with binder indices and BINDER INDEX ECONOMY offer some insight into the pattern of processing results reported above. When readers encounter *they*, it is given an enriched representation – seeking [-SG] and [-FEM,-MASC].²⁰ If the element it retrieves is singular, this will lead to a clash. Hence the across-the-board spillover latency we found in both Experiments 1b and 2b. The crucial interaction with quantification comes with the gender. All instances of *they* will be enriched and thus bear [-FEM,-MASC] features that clash with gendered antecedents (at least for conservative English users). The difference between referential and quantificational antecedents is the following. A referential DP, which carries gender features, serves as an antecedent, while a QP is not in any technical sense an antecedent. Rather, the antecedent is the binder index, and it does not bear gender features due to BINDER INDEX ECONOMY. When pronouns retrieve referential DPs, they cannot avoid getting all the DP features. Pronouns do not retrieve quantifier phrases, only their associated binder indices, and *these* – for the principled reasons we sketched above – are very naturally underspecified for gender.

Turning to number, this approach also provides an understanding of the asymmetry between number and gender that we found in the online reading results. At first blush, we might have expected that BINDER INDEX ECONOMY would successfully allow the suppression of number on the binder index just as it does gender features. But we did find, even in non-gendered bound variable cases, a cost on the spillover region for singular *they*. This might again follow from the nature of the English verbal agreement system, which expones number but not gender. Number is thus a more likely candidate than gender to be expressed on a verbal head and so the processor might put it on the binder initially. This would lead to a presupposition violation when *they* is encountered downstream. A speaker's exposure to the contrasts made in the overt verbal agreement system thus can have effects of the likelihood of encoding features on binder indices. However, number could still be removed from the

singular *they* with gendered referential antecedents. Of course, it is possible that readers *do* pursue a bound interpretation in the referential conditions on some number of trials, and this could be the source of acceptable uses of singular *they* in those conditions (recall that across the board, we found that singular *they* was rated high even with referential antecedents).

²⁰Recall that we already have found reasons to rule out the underspecification approach.

binder index eventually due to BINDER INDEX ECONOMY, not affecting offline reflective judgments.²¹

7. CONCLUSION

We have confirmed that bound-variable singular *they* has a processing advantage over referential singular *they*. In offline acceptability judgments, bound *they* is in fact preferred over a singular gendered alternative when the antecedent is non-gendered and just as acceptable when the antecedent is gendered. Referential singular *they* was never preferred to its singular gendered alternative, regardless of antecedent type. The reading time results also revealed an advantage for bound *they*, but only with gendered antecedents.

A secondary and revealing finding was a consistent reading-time cost for singular *they* in the region following the pronoun across all antecedent manipulations. We suggested that this reflects a weak number effect, one that is easily overcome in the offline tasks. This finding shows that even the most favourable environment for singular *they* – as a variable bound by a non-gendered antecedent – still poses a cost. We suggested that this temporary sensitivity to number might follow from the English verbal agreement system which expones number, but not gender.

We modeled the overall advantage for bound versus referential *they* in terms of a grammatical preference to minimize the features on binder indices, a move that distinguishes bound variables from referential pronouns. This move has implications for the processing mechanisms of antecedent retrieval, suggesting that not all antecedent types are retrieved in the same way.

SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit <https://doi.org/10.1017/cnj.2022.30>.

REFERENCES

- Ackerman, Lauren. 2018. Processing singular *they* with generic and specific antecedents. Presented at AMLaP (Architectures and Mechanisms for Language Processing), Berlin.
- Ackerman, Lauren. 2019. Syntactic and cognitive issues in investigating gendered coreference. *Glossa* 4(1): 117.
- Ackerman, Lauren, Nick Riches, and Joel Wallenberg. 2018. *Coreference dependency formation is modulated by experience with variation of human gender*. Presented to the annual meeting of the Linguistic Society of America, Salt Lake City.

²¹A reviewer observes that some speakers of an African American English dialect do not have third person singular agreement in their verbal inflection and that our analysis predicts that such speakers should not exhibit any difficulty in processing bound singular *they* with either non-gendered or gendered quantificational antecedents. We thank the reviewer for pointing this out to us.

- Adger, David, and Gillian Ramchand. 2005. Merge and move: *Wh*-dependencies revisited. *Linguistic Inquiry* 36(2): 161–193.
- Balhorn, Mark. 2004. The rise of epicene *they*. *Journal of English Linguistics* 32(2): 79–104.
- Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3): 255–278.
- Bates, Douglas M. 2005. Fitting linear mixed models in R: Using the lme4 package. *R News: The Newsletter of the R Project* 5: 27–30.
- Bjorkman, Bronwyn M. 2017. Singular *they* and the syntactic representation of gender in English. *Glossa* 2(1): 80. 1–13.
- Bodine, Ann. 1975. Androcentrism in prescriptive grammar: Singular ‘they’, sex-indefinite ‘he’, and ‘he or she’. *Language in Society* 4(2): 129–146.
- Camilliere, Sadie, Amanda Izes, Olivia Leventhal, and Daniel Grodner. 2019. Pragmatic and grammatical factors that license singular *they*. Paper presented at XPrag 2019, University of Edinburgh.
- Carminati, Maria Nella, Lyn Frazier, and Keith Rayner. 2002. Bound Variables and C-Command. *Journal of Semantics* 19(1): 1–34.
- Carreiras, Manuel, Alan Garnham, Jane Oakhill, and Kate Cain. 1996. The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *The Quarterly Journal of Experimental Psychology Section A* 49(3): 639–663.
- Conrod, Kirby. 2019. *Pronouns raising and emerging*. Doctoral dissertation, University of Washington.
- Cunnings, Ian, Clare Patterson, and Claudia Felser. 2014. Variable binding and coreference in sentence comprehension: Evidence from eye movements. *Journal of Memory and Language* 71(1): 39–56.
- Déchaine, Rose-Marie, and Martina Wiltschko. 2002. Decomposing pronouns. *Linguistic Inquiry* 33(3): 409–442.
- den Dikken, Marcel. 2006. *Relators and linkers: The syntax of predication, predicate inversion, and copulas*. Cambridge, MA: MIT Press.
- Doherty, Alice, and Kathy Conklin. 2017. How gender-expectancy affects the processing of “them”. *The Quarterly Journal of Experimental Psychology* 70(4): 718–735.
- Drummond, Alex. 2013. *Ibex farm*. <<http://spellout.net/ibexfarm/>>
- Ferreira, Fernanda, Karl G. D. Bailey, and Vittoria Ferraro. 2002. Good-enough representations in language comprehension. *Current Directions in Psychological Science* 11(1): 11–15.
- Ferreira, Fernanda, and Charles Clifton. 1986. The independence of syntactic processing. *Journal of Memory and Language* 25(3): 348–368.
- Filik, Ruth, Anthony J. Sandford, and H. Leuthold. 2008. Processing pronouns without antecedents. *Journal of Cognitive Neuroscience* 20(7): 1315–1326.
- Fodor, Janet Dean, and Ivan A. Sag. 1982. Referential and quantificational indefinites. *Linguistics and Philosophy* 5(3): 355–398.
- Foertsch, Julie, and Morton Ann Gernsbacher. 1997. In search of gender neutrality: Is singular *they* a cognitively efficient substitute for generic *he*? *Psychological Science* 8(2): 106–111.
- Frazier, Lyn and Charles Clifton. 2000. On bound variable interpretations: The LF-only hypothesis. *Journal of Psycholinguistic Research* 29(2): 125–140.
- Grange, James A. 2015. trimr: Response Time Trimming in R. <<https://cran.r-project.org/web/packages/trimr/vignettes/trimr-vignette.html>>
- Grodzinsky, Yosef, and Tanya Reinhart. 1993. The innateness of binding and coreference. *Linguistic Inquiry* 24(1): 69–102.

- Heim, Irene. 1982. *The semantics of definite and indefinite Noun Phrases. Doctoral dissertation, University of Massachusetts*, Amherst.
- Heim, Irene and Angelika Kratzer. 1998. *Semantics in Generative Grammar*. Malden, MA: Blackwell.
- Heycock, Caroline. 1992. Layers of predication and the syntax of the copula. *Belgian Journal of Linguistics* 7: 95–123.
- Just, Marcel Adam, Patricia A. Carpenter, and Jacqueline D. Woolley. 1982. Paradigms and processing in reading comprehension. *Journal of Experimental Psychology: General* 111(2): 228–238.
- Kamp, Hans. 1981. A theory of truth and semantic representation. In *Formal methods in the study of language*, ed. Jeroen Groenendijk, Theo Janssen, and Martin Stokhof, 277–322. Amsterdam: Mathematical Center.
- Konnely, Lex, and Elizabeth Cowper. 2020. Gender diversity and morphosyntax: An account of singular *they*. *Glossa: A Journal of General Linguistics* 5(1): 40. 1–19. DOI: <http://doi.org/10.5334/gjgl.1000>.
- Koornneef, Arnout, Sergey Avrutin, Frank Wijnen, and Eric Reuland. 2011. Tracking the preference for bound-variable dependencies in ambiguous ellipses and only-structures. In *Experiments at the interfaces 2*, ed. Jeffrey Runner, 67–100. Leiden: Brill.
- Koornneef, Arnout Willem. 2008. *Eye-catching anaphora*. Utrecht: LOT.
- Kratzer, Angelika. 1998. Scope or pseudo-scope? Are there wide-scope indefinites? In *Events in grammar*, ed. Susan Rothstein, 163–196. Dordrecht: Kluwer.
- Kratzer, Angelika. 2009. Making a pronoun: Fake indexicals as windows into the properties of pronouns. *Linguistic Inquiry* 40(2): 187–237.
- Kreiner, Hamutal, Patrick Sturt, and Simon Garrod. 2008. Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language* 58(2): 239–261.
- Kuznetsova, Alexandra, Per Bruun Brockho, and Rune Haubo Bojesen Christensen. 2014. lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer Objects of lme4 Package). R package version 2.0-11.
- Matthewson, Lisa. 1999. On the interpretation of wide-scope indefinites. *Natural Language Semantics* 7(1): 79–134.
- Moro, Andrea. 1997. *The raising of predicates: Predicative noun phrases and the theory of clause structure*. Cambridge: Cambridge University Press.
- Moulton, Keir, Chung-hye Han, Trevor Block, Holly Gendron, and Sander Nederveen. 2020. Singular *they* in context. *Glossa* 5(1): 122. 1–13. DOI: <https://doi.org/10.5334/gjgl.1012>.
- Moxey, Linda M., Anthony J. Sanford, Patrick Sturt, and Lorna I. Morrow. 2004. Constraints on the formation of plural reference objects: The influence of role, conjunction, and type of description. *Journal of Memory and Language* 51(3): 346–364.
- Nouwen, Rick. 2003. Complement anaphora and interpretation. *Journal of Semantics* 20(1): 73–113.
- Oosterhout, Lee, and Linda A. Mobley. 1995. Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language* 34(6): 739–773.
- Paterson, Kevin B., Ruth Filik, and Linda M. Moxey. 2009. Quantifiers and discourse processing. *Language and Linguistics Compass* 3(6): 1390–1402.
- Phillips, Colin. 2006. The real-time status of island phenomena. *Language* 82(4): 795–823.
- R Development Core Team. 2020. R: A language and environment for statistical computing. <<http://www.R-project.org>>
- Reinhart, Tanya. 1997. Quantifier scope: How labor is divided between QR and choice functions. *Linguistics and Philosophy* 20(4): 335–397.

- Robitzsch, Alexander. 2020. Why ordinal variables can (almost) always be treated as continuous variables: Clarifying assumptions of robust continuous and ordinal factor analysis estimation methods. *Frontiers in Education* 5. DOI: <https://doi.org/10.3389/educ.2020.589965>.
- Rullmann, Hotze. 2003. Bound-variable pronouns and the semantics of number. In *Proceedings of WECOL 2002*, ed. Brian Agbayani, Paivi Koskinen, and Vida Samiian, 243–254.
- Sanford, Anthony J., and Ruth Filik. 2007. “They” as a gender-unspecified singular pronoun: Eye tracking reveals a processing cost. *The Quarterly Journal of Experimental Psychology* 60(2): 171–178.
- Sanford, Anthony J., Ruth Filik, Catherine Emmott, and Lorna Morrow. 2008. Short article: They’re digging up the road again: The processing cost of institutional *they*. *Quarterly Journal of Experimental Psychology* 61(3): 372–380.
- Sauerland, Uli, Jan Anderssen, and Kazuko Yatsushiro. 2005. The plural is semantically unmarked. In *Linguistic evidence: Empirical, theoretical, and computational perspectives*, ed. Stephan Kepser and Marga Reis, 413–434. Berlin: Mouton de Gruyter.
- Schu'tze, Carson, and Jon Sprouse. 2014. Judgment data. In *Research methods in linguistics*, ed. Robert J. Podesva and Devyani Sharma, 27–50. Cambridge: Cambridge University Press.
- Schwarzschild, Roger. 2002. Singleton indefinites. *Journal of Semantics* 19 (3): 289–314.
- Sudo, Yasutada. 2012. *On the semantics of phi features on pronouns*. Doctoral dissertation, Massachusetts Institute of Technology.
- Sudo, Yasutada. 2014. Dependent plural pronouns with Skolemized choice functions. *Natural Language Semantics* 22(3): 265–297.
- Trueswell, John C., and Michael K. Tanenhaus. 1994. Semantic influence on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language* 33(3): 285–318.
- Winter, Yoad. 1997. Choice functions and the scopal semantics of indefinites. *Linguistics and Philosophy* 20(4): 399–467.