TRACKING REFERENTIAL ADAPTATION TO NONBINARY THEY USING A MOUSE-TRACKING PARADIGM

Yining Ye

A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology & Neuroscience.

Chapel Hill
2023

Approved by:

Jennifer Arnold

Kelly Giovanello

Peter Gordon

ABSTRACT

Yining Ye: Tracking Referential Adaptation To Nonbinary They in A Mouse-Tracking Paradigm
(Under the direction of Jennifer E. Arnold)

In referential adaptation, people adapt to the relationships between a pronoun and its antecedent (Johnson & Arnold, 2022). Most evidence comes from offline tasks testing how exposure influences resolution of subject and nonsubject referents. Arnold et al. (2023) used mouse-tracking to examine processing of singular vs. plural *they* pronouns. The singular *they* elicits a processing difficulty. Based on these findings, the current study tests 1) the sensitivity of mouse-tracking to processing of subject and nonsubject references and 2) adaptation to singular *they*. Participants listened to stories of two characters doing an activity followed by a pronoun, which is disambiguated by a target object placed under one of the characters. Participants clicked on the target object. Mouse movements and RTs were analyzed. Two pilots tested the processing of subject and nonsubject interpretations but failed to replicate the subject bias. The main experiment exposed participants to singular or plural *they* and found adaptation effects.

ACKNOWLEDGEMENTS

First of all, I would like to sincerely express my gratitude to my advisor, Jennifer E. Arnold for all her guidance and support for the development and completion of this project and the paper. I would also like to express my gratitude to my committee members for their feedback and suggestions. Also, I would like to thank my lab members and all undergraduate research assistants helping with this project. And I really appreciate my friends and families for their emotional support these years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATONS

MAD	Maximum Absolute Deviation
AUC	Area Under the Curve
X-flips	Directional reversals along the horizontal axis

INTRODUCTION

It has been well established that people adapt to various statistical regularities in language. In language comprehension, adaptation occurs when people's language representations change in response to the linguistic inputs they recently encounter (Kaan & Chun, 2018; Fine et al., 2010; Prasad & Linzen, 2021). With exposure, people may use those inputs as a guide when they later encounter similar but ambiguous structures. They may also become faster at processing similar inputs. This adaptation phenomenon has been extensively studied at multiple levels of language. For example, at the syllabic level, infants could learn artificial syllable sequences and use this information to guide future speech segmentation (Saffran et al., 1998). At the lexical level, studies found that high-frequency words (e.g., baby) tend to be recognized and processed more quickly than words (e.g., eczema) that are infrequent in the language (e.g., Brysbaert et al., 2018; Monsell et al., 1989). There is also extensive evidence presented for syntactic adaptation, in which exposure to rare syntactic structures facilitates people with subsequent processing of similar structures (e.g., Fine & Jaeger, 2013).

More recently, a line of work has started to look at adaptation at the discourse-level. Specifically, studies tested referential adaptation, in which people adapt to the relationship between a pronoun and its antecedent, which is the corresponding referent in the preceding discourse (Kaiser, 2009). Studies have provided evidence for people exhibiting referential adaptation to relationships between a third-person pronoun (e.g., he/she) and its antecedent. Exposure to these pronoun-antecedent relationships were hypothesized to result in a memory for the abstract structure, which can be characterized in terms of the antecedent's semantic or grammatical role. For example, a pronoun can refer to the subject referent who is the first-mentioned in the previous sentence, or to the nonsubject referent who is the second-mentioned. In a sentence "*Ana went to the library with Will and then she borrowed a book*" where the pronoun *she* refers to Ana, the relationship can be seen as an instance of a referential structure where a pronoun refers to an antecedent in the subject position of the previous sentence. Most of the studies used an offline pronoun comprehension task to measure people's pronoun interpretation preferences, and found that exposure to a reference pattern drives people to follow the pattern while interpreting ambiguous pronouns (e.g., Johnson & Arnold, 2022). Therefore, exposure influences people's behavioral outcomes in pronoun comprehension. However, it remains unclear whether exposure also affects online processing. Therefore, the current study starts with the question about online referential adaptation. Do people adapt to referential structures by getting faster in resolving ambiguous pronouns online?

In this investigation, we aimed to use mouse-tracking as our online task. During the pandemic when in-person data collection was restricted, mouse-tracking was a convenient way to collect online data as it can be implemented via web platforms like PC-Ibex. Before testing referential adaptation, we would like to verify whether mouse-tracking is sensitive enough to reveal the differential processing of the subject vs. nonsubject references. Referential processing is constrained by a series of factors that determine a referent's prominence in the discourse, including the referent's grammatical positions. Specifically, it is well-known that people are usually biased towards the subject of the previous sentence when interpreting an ambiguous pronoun (e.g., Arnold et al., 2000; Gordon & Scearce, 1995). Therefore, people tend to have more subject than nonsubject interpretations of pronouns in comprehension tasks. This difference is expected to be present during online processing as well, such that people are faster at processing a referential relationship between a pronoun and a subject referent, than one between a pronoun and a nonsubject referent.

We expected to observe these differences in processing because the mouse-tracking paradigm developed by Arnold et al. (2023) is sensitive to the differential processing of a different type of pronoun ambiguity, which is between singular vs. plural *they* pronouns. The singular *they* use for nonbinary individuals has gained an increasing distribution in the language, but it's still a relatively new usage compared to the common plural *they* used to refer to multiple or a group of people. Arnold et al.'s (2023) mouse-tracking paradigm revealed a processing difficulty elicited by the singular use of *they* in competition with a bias toward the common plural use of *they*. This sensitivity to ambiguous referential processing shows that the mouse-tracking paradigm is likely a suitable method to test referential processing of subject- and nonsubject-pronoun structures.

Primarily, the current study investigates the influence of exposure on online processing of referential adaptation. Based upon the above findings, we aimed to use Arnold et al.'s (2023) mouse-tracking paradigm in the current study to answer two questions. First, we ask whether the mouse-tracking paradigm is sensitive to the processing of subject vs. nonsubject references. However, our pilot studies using this paradigm did not reveal any reliable subject bias. As a result, we decided to test adaptation using the singular vs. plural *they* ambiguity. Based on the behavioral findings on referential adaptation, our second question asks whether people can also adapt to interpretations of different pronouns like singular *they* as they adapt to the interpretations of third-person singular pronouns.

Referential Adaptation

Referential adaptation is defined as a process, in which frequent exposure to a relationship between a pronoun and its antecedent guides future pronoun comprehension (Johnson & Arnold, 2022). Do people form any representations about these implicit referential relationships? Initial evidence does demonstrate comprehenders' ability to mentally represent and adapt to referential structure through exposure. An earlier study by Kaiser (2009) used a priming technique to test whether a brief exposure to a referential pattern influences subsequent processing of the same configuration. Specifically, they gave participants a referential structure, in which the pronoun referred to either the subject or object character (e.g., William swooked Betty and Kevin brucked her/him). The verbs were all made-up words to avoid other semantic effects in pronoun comprehension. Following this exposure sentence, participants were given a critical sentence (e.g., Stephen tulvered Peter and Diane churbited him) followed by a two-choice question (e.g., Diane churbited_____. Stephen? Peter?) to indicate their pronoun interpretations. The results showed that participants were more likely to adopt the prime structure (e.g., pronoun refers to the subject) in responses of pronoun interpretations. This early evidence suggests people do represent the implicit link between the pronoun and the corresponding referent. Similar results were found in L2 English speakers by Contemori (2021) who demonstrated that the primary subject bias could be modulated by exposure to interpretation biases towards the nonsubject character. Following each priming sentence that had a pronoun referring to the nonsubject, participants were more likely to adopt nonsubject reference pattern when interpreting ambiguous pronouns. These studies provided important evidence for comprehenders' ability to form the abstract relationship between a pronoun and its antecedent, and that a single exposure is enough to drive people's subsequent pronoun interpretations.

Kaiser (2009) and Contemori (2021) provided evidence for immediate priming of referential structures because they tested how one exposure item affected the immediately following test item. Taking a closer look at adaptation, Goodrich Smith et al. (2019) provided preliminary evidence that children as young as 5 years old can learn pronoun interpretation biases from exposure. The study exposed participants to narratives that consistently displayed the pattern of pronouns referring to the first-mentioned characters. Following exposure, children were tested on comprehension questions to indicate their pronoun interpretation preferences. The training by exposure stories lasted across five days for the first-mentioned bias to be established. Finally, compared to the baseline interpretation preference measured in a pre-test before exposure, the children's assignment of the pronouns to the first-mentioned character increased from 48% to 69%, demonstrating that exposure to a referential pattern modulated children's pronoun interpretation biases.

The Referential Adaptation Paradigm

The above studies provided preliminary evidence for people exhibiting adaptation to referential structures, using a pronoun comprehension task to measure participants' pronoun interpretation preferences following exposure. While the effect of exposure found in previous studies was small, a more recent study by Johnson and Arnold (2022) has captured more robust effects. They used a referential adaptation paradigm to test whether comprehenders adapt to the most frequent referential structure in a local context. The task first exposed participants to a series of unambiguous referential structures with pronouns that either always referred to the subject ("Ana painted the wall with Will. She…") or to the nonsubject ("Ana went to the library with Will. He…").

Following the exposure to a particular referential pattern, participants were then tested on referential structures that contained ambiguous pronouns, which assessed participants' own interpretation biases. There were 40 exposure sentences that contained unambiguous pronouns in order to establish an assigned referential pattern; 20 occurred at the beginning of the experiment and the others were intermixed with critical trials throughout the experiment. Following each item, there were two comprehension questions, one asking about the interpretation of the pronoun (e.g., "Did Liz paint the wall?") and one asking about the story content to assess whether participants paid attention to the task. Therefore, the primary manipulation of this experiment is the types of exposure patterns (pronouns refer to the subject vs. pronouns refer to the nonsubject).

Results of pronoun interpretations showed a main effect of exposure, such that participants gave higher proportion of subject interpretations for pronouns following subject-reference patterns or higher proportion of nonsubject interpretations for pronouns following nonsubject-reference pattern. The referential adaptation paradigm successfully captures comprehenders' tendency to follow the exposure pattern of referential structures in pronoun interpretations. Based on this study, we integrated our online task with this referential adaptation paradigm to manipulate the particular type of referential structure participants were exposed to.

Subject Bias in Referential Processing

The learning of referential structures can be influenced by multiple constraints, including syntactic, semantic and pragmatic cues, that elicit different interpretation biases for an ambiguous pronoun (e.g., Arnold, 1998; Arnold et al., 2000; Kehler et al., 2008; Rohde & Kehler, 2014). People's change of interpretation biases by adaptation to the exposure referential structure may depend on their preexisting preferences of pronoun interpretations. One of the well-known syntactic constraints that drives ambiguous pronoun resolution is the subject bias, which is a general tendency to assign the subject antecedent in the previous sentence to ambiguous pronoun (e.g., Gordon & Scearce, 1995; Kehler & Rohde, 2019; Langlois & Arnold, 2020). For example, in a sentence “Ana washed the towel with Liz and then she dried the plates” where the ambiguous pronoun can refer to either Ana or Liz, people generally expect the subject Ana to be mentioned next and therefore adopt a “subject-assignment” strategy while resolving the pronoun.

This subject bias provides a starting point to examine how people adapt to different referential structures and how local exposure to a structure changes these predominant biases. In Johnson and Arnold (2022)’s adaptation study, one of the constructions being tested was called the joint-action construction, in which the verb identifies a subject and a nonsubject following the preposition “with” (e.g., Liz went to a supermarket with Ana. She bought some bread.). This construction has been extensively tested and is known to elicit a strong bias towards the subject (e.g., Arnold et al., 2018). This subject bias was found to be modulated by exposure to a series of nonsubject reference sentences (e.g., Liz went to a supermarket with Matt. He bought some bread.), in which the pronouns always refer to the nonsubject. Specifically, Johnson and Arnold’s (2022) results showed that participants’ subject interpretations of ambiguous pronouns were significantly reduced following exposure to nonsubject reference patterns.

This finding demonstrates the subject- vs. nonsubject-pronoun structures as a suitable case to study the influence of exposure on referential adaptation. The behavioral results show that participants tend to follow the exposure pattern in their pronoun interpretations indicated by a discrete decision to choose either the subject or nonsubject. What these offline behavioral changes implied is the resulting changes in one’s representations of what type of referential structure is more probable. However, it will be difficult to use response tendency to make inferences about how people get to the final pronoun interpretation and whether they experience any competition during processing. Offline measures such as accuracy are used to observe the behavioral outcome, while online measures such as eye or mouse movements are methods for assessing the characteristics of ongoing psychological processes (Kieslich et al., 2019; Freeman et al., 2010).

The current study focuses on how exposure affects online processing of referential adaptation. Before testing adaptation, the first objective of the study is to investigate the sensitivity of mouse-tracking to the differential processing of subject- and nonsubject-pronoun referential structures. We aimed to replicate the behavioral findings (e.g., the subject bias) in previous studies. Specifically, it is expected that mouse-tracking would reveal an easier and faster processing of subject-pronoun structures compared to nonsubject-pronoun structures. However, this pattern was not observed in our mouse-tracking paradigm, so we did not follow up to further test adaptation to subject- and nonsubject-references. In contrast to our finding, Arnold et al. (2023) showed that the mouse-tracking paradigm is sensitive to the ambiguity of singular vs. plural *they*. Based on this study, we then used the paradigm to examine the effect of exposure on online referential adaptation to singular vs. plural *they*.

Do people also adapt to nonbinary they?

Based on the findings on referential adaptation, it's clear that exposure to subject-third-person-pronoun relationships and exposure to nonsubject-third-person-pronoun relationships lead to increases in subject- and nonsubject-bias interpretations respectively. The current study extends the question to comprehension of the pronoun *they*.

The prescriptive rule of English defines *they* as a plural pronoun that co-refers with multiple or a group of people (e.g., “The beach was full of visitors. They were waiting for the sunset.”). However, *they* has long been used differently in various contexts. For example, it has been widely common that *they* can be used as a third-person generic pronoun to refer to singular, nonspecific or gender-unknown antecedents (e.g., “They are a friend of mine”) in epicene contexts (e.g., Noll et al., 2018). More recently, there is an increasing trend to use *they* as a singular pronoun for nonbinary individuals (e.g., “Alex went to the store. They bought a sandwich.”). The expanding distribution of this relatively new use of *they* in English has been of great interest to researchers (e.g., Conrod, 2020) but also raised great controversies among the discussions of acceptability and grammaticality of singular *they*. Bjorkmam (2017) pointed out that speakers and listeners distinctively varied in their use of singular *they*. For example, “non-innovative” users would consider *they* as grammatically incompatible with explicit gender marking and be resistant to use *they* in these contexts (e.g., “Sophia went to the store because they needed apples.”). Bradley et al. (2019) also found that people would reject the use of nonbinary *they* based on gender ideology.

Nevertheless, the use of nonbinary singular *they* is rapidly increasing. More people declare the use of this pronoun when others are referring to them and this requests other speakers to also understand and use nonbinary *they*. Then what drives the increasing frequency of nonbinary *they* in the language? One possibility is that there is a strong social trend to avoid misgendering, promote respect for and politeness to individuals’ pronoun choices. This increasing awareness makes nonbinary they socially salient (Conrod, 2020). At the same time, learning to understand and use a relatively new and low-frequency pronoun requires practice and contexts that support the learning process. Therefore, it’s highly likely that people are adapting to the nonbinary use of *they* through relevant exposure and language experience. Then the question is whether people can adapt to nonbinary *they* as they adapt to the subject-pronoun and nonsubject-pronoun relationships. If people do adapt, it's expected that adaptation would also modulate the differences in comprehension of singular versus plural *they*.

Previous studies have found that singular and plural *they* are comprehended differently. They show that singular *they* elicits a processing cost that is not observed in interpretation of other common and frequent pronouns like plural *they*. For example, Leventhal et al. (2020) conducted an ERP study to test whether nonbinary *they* is processed differently from other singular pronouns (e.g., *he*, *she*). During ERP recording, 32 participants who were familiar with nonbinary gender identities read sentences with pronouns referring to either gender- or number-match and -mismatch antecedents, shown as example (1) below. After reading each sentence, participants were asked what gender they stereotypically associated with the subject antecedent in that sentence.

(1) ***Gender-matching singular pronoun***: Lillian had just gotten back from a vacation, so she felt exhausted.

Gender-mismatching singular pronoun: Lillian had just gotten back from a vacation, so he felt exhausted.

Number-matching plural pronoun: Lillian and Paul had just gotten back from a vacation, so they felt exhausted.

Number-mismatching plural pronoun: Lillian had just gotten back from a vacation, so they felt exhausted.

Results show that sentences of both gender-mismatch and number-mismatch conditions elicited a P600 effect, which indicated a grammatical processing difficulty in comprehending the mismatching pronouns. Therefore, even for people who are familiar with nonbinary gender identities, they experience processing difficulty when comprehending singular *they*.

However, Arnold et al. (2021) showed that comprehension of singular *they* could be facilitated if there is explicit introduction about the pronoun use. In the study, participants were introduced to three characters (Alex, Liz and Will) who use *they*, *he*, *she* pronouns respectively. Participants were assigned to conditions, in which they were given either an explicit or implicit introduction. The explicit condition emphasized the pronoun each character used, while the implicit condition only gave the character names. Following the introduction, participants read a two-sentence story shown as example (2) below. Experiment 1 had stories either about Alex only or about Alex with another character with Alex always being the first-mentioned, while Experiments 2 and 3 had all two-character stories and manipulated whether Alex was first- or second-mentioned. The task was to click on character pictures to answer a pronoun interpretation question (e.g., who did…?) and to answer a multiple-choice content question. The pictures were either one-character picture or two-character picture, representing a singular and a plural interpretation of the pronoun respectively.

(2) ***One-character story***: Alex went running. They fell down.

Two-character story (Alex as first-mentioned): Alex went running with Liz. They fell down.

Two-character story (Alex as second-mentioned): Liz went running with Alex. They fell down.

Pronoun comprehension task: Who fell down?

Content question: What did Alex do? (Went running/Had coffee at Starbucks)

Based on the pronoun comprehension task results, the study found that the explicit pronoun introduction led to higher rate of singular interpretations of the pronoun *they* in the two-character stories that people were more likely to click on the picture of Alex alone. Also, the singular interpretation was also facilitated by order-of-mention. In Experiments 2 and 3, people gave more singular interpretations when Alex was the first-mentioned. Overall, the study used an offline task, revealing that the singular interpretations of *they* could be promoted by explicitly introducing a character's personal pronouns.

The Mouse-Tracking Paradigm

Following Arnold et al. (2021)'s design, Arnold et al. (2023) conducted a mouse-tracking study to examine the differences in online processing of singular versus plural *they*. The study also explicitly introduced that the character Alex used *they* pronouns and another two characters Liz and Will used *she* and *he* pronouns. In the mouse-tracking paradigm, participants read three-sentence stories shown as example (3) below and were presented with three pictures. The critical stories always had Alex and one of the other two characters Liz or Will. Alex was manipulated to be either first- or second-mentioned in the story. The corresponding pictures were then Alex, the other character or both together. Under each picture were two objects, only one was mentioned in the spoken story. Participants were asked to click on the mentioned object, which was manipulated to be placed under the picture of either Alex alone or both characters together. These object locations under different character pictures indicated a singular and a plural interpretation of the *they* pronoun. During the process, participants' reaction times and mouse-trajectories were recorded for further analysis.

(3) ***Alex as first-mentioned***: Alex and Liz were cleaning up after a dinner party. Alex snatched a towel from Liz. Then they dried the plates.

Alex as second-mentioned: Liz and Alex were cleaning up after a dinner party. Liz handed a towel to Alex. Then they dried the plates.

The study compared the differences in online measures between the two object location conditions and examined the secondary effect of Alex's order-of-mention on how people reacted. Reaction times were longer in the condition where the target object was placed under Alex, which means that people were slower in clicking the object that indicated a singular interpretation. One of the mouse-tracking measures (x-flips_after_pronoun, the directional changes of the mouse alone the horizontal axis on the screen), which measured the degree of competition among available choice candidates during processing, were higher in the condition where the target object was placed under Alex. This suggested that people seemed to experience a stronger competition when they saw the target object was under a singular person. However, these effects were modulated by order-of-mention, such that the difference between conditions was most promising when Alex was the second-mentioned, possibly because people had to overcome both the dominant plural interpretation of *they* and the first-mention bias. Overall, this study provided important evidence for online processing of singular *they*, that people with explicit pronoun introduction still experience difficulty processing this low-frequency pronoun.

Based on the above findings, it's clear that singular *they* is more difficult to process than other singular pronouns like *he* and *she*, and also the plural *they*. Although the explicit introduction by itself alone increases the saliency of the pronoun status, it is not enough to eliminate the processing cost of singular *they* to make it compete with the dominant plural *they*. Therefore, the current study examines whether exposure could modulate the differences in processing the different usages of *they*. It was expected that adaptation to either singular or plural *they* would inhibit the processing of the other pronoun use. It's even possible that strong adaptation might even overturn the dominance status of plural *they* within the experiment that people might experience difficulty processing plural *they* instead. This finding also demonstrated that this mouse-tracking paradigm might be used to test referential processing of subject and nonsubject references.

The current study

The current study builds on the offline findings from Johnson and Arnold (2022) and online findings from Arnold et al. (2023), integrating the referential adaptation paradigm with the mouse-tracking paradigm to examine the effect of exposure on referential adaptation.

The findings of Arnold et al. (2023) provided an exploratory paradigm of using mouse-tracking to test online referential processing of *they* pronouns and show that at least some of mouse-tracking measures including reaction times and x-flips are sensitive enough to reveal differential processing patterns of singular versus plural *they*. Following this study, one objective of the current study was to test whether this paradigm could be used to test online referential processing of subject- and nonsubject-pronoun structures that were tested the most in referential adaptation. Most studies in this investigation used offline tasks like pronoun comprehension questions. Therefore, the current study tested whether the mouse-tracking paradigm can reveal any differences in processing referential structures of pronouns referring to subject or nonsubject antecedents, corresponding to the behavioral tendences that people are biased towards the subject more than the nonsubject antecedent (e.g., Johnson & Arnold, 2022). The study started with two pilot studies (on-web vs. in-lab) to test whether mouse trajectories and reaction times could replicate any differences in pronoun comprehension in structures containing joint-action verbs, which are well-known to elicit a strong subject bias. The pilot used a modified version of the paradigm used in Arnold et al. (2023). Participants were introduced to four characters (Liz, Ana, Matt and Will) and listened to stories. For each story, participants were presented with two pictures of each individual character, and then listened to stories consisting of three sentences. The last sentence contained a pronoun referring to either the subject or the nonsubject antecedent. Under each character picture there were two objects. The target object was manipulated to be placed under the subject or the nonsubject character, and participants' reaction times and mouse trajectories were recorded for further analysis. However, neither pilot revealed any promising sensitivity of mouse-tracking to the differential processing of subject- and non-subject-pronoun structures. Possible explanations are provided in the general discussion.

Further, following Arnold et al. (2023)'s design, the study served as an exploratory investigation of online processing in referential adaptation to the use of singular *they*, as no evidence has been provided to answer the question about how exposure influences online processing of singular *they*. So the other objective of the study is to track referential adaptation to singular *they* in a mouse tracking paradigm. With introduction to the characters and their pronoun uses (*Alex-they, Liz-she, Will-he*), participants were exposed to a series of exposure stories containing pronouns that either required a singular or a plural interpretation of *they*, disambiguated by the target object location. With exposure, participants were tested on critical stories of Alex with another character to see how they differed in processing the singular versus plural interpretations of *they*.

PILOT 1

The pilot study followed the mouse-tracking paradigm developed by Arnold et al. (2023) and examined whether mouse-tracking is sensitive to referential processing in general. Specifically, the pilots tested whether it reveals any differences in comprehension of subject-pronoun and nonsubject-pronoun relationships.

Methods

Participants

59 Amazon Mechanical Turk workers participated in exchange for $2.50. The Mturk HIT was only available to native English speakers from the US, Australian and England. The participants were required to have 1000 previous HITs with at least 95% acceptance rate. All participants reported learning English before age 6 and a native level of proficiency. Out of the 59 participants who completed the study, 28 were excluded for low accuracy, leaving 31 participants for further analysis. The 28 Participants were excluded for achieving lower than 85% accuracy across all items. As the task was easy and required low-level cognitive demand, participants with accuracy below the required level were considered inattentive or possibly rushing through the experiment without really listening to the stories.

Materials and Design

The experiment was implemented on web via the PC-Ibex platform, which included background demographic questions and the main story task. Demographic questions asked about age, sex, gender, ethnicity/race, language experience, language proficiency and report of language disorders. To ensure the spoken stimuli would be received appropriately, participants were given an audio test (“The number is seven.”) to click on one of the numbers from 1 to 7. The experiment would terminate if participants did not choose the mentioned number.

For the main task, the setup of the experiment is illustrated in Figure 1. The basic paradigm introduces 4 characters (Ana, Liz, Will and Matt) and a series of short stories. Each story consists of a context sentence introducing 2 characters and a pronoun sentence with a pronoun referring to either the subject or the nonsubject character, as shown in Table 1. For each story, two pictures of the mentioned characters were placed either left or right side of the screen. Two pairs of objects were placed under the characters respectively. The visual objects could be potentially fit into the sentence but only one target object was mentioned in the spoken story. The target object was manipulated to be placed under either the subject or non-subject character to indicate a subject- or non-subject interpretation and thus disambiguate the pronoun.

The 16 critical stories all included two same-gender characters so the pronouns were all ambiguous and could only be disambiguated by target object location. The 26 filler stories were all unambiguous stories where the pronoun could be disambiguated by either the gender information in the spoken stimuli or the visual target location. To reinforce the subject bias, all fillers had pronouns that referred to the subject characters. Before the main task, 2 unambiguous stories as practice familiarized participants with the task. All stimuli are shown in Appendix 2.

Overall, the experiment had one within-subjects manipulation for the critical items, which had the target objects to be placed under pictures of either the subject character or the nonsubject character. The target location was counterbalanced within each item, resulting in 2 lists. Across all items, the locations of the target characters and target objects were counterbalanced so that each target character and object appeared on the left and right side of the screen the same number of times.

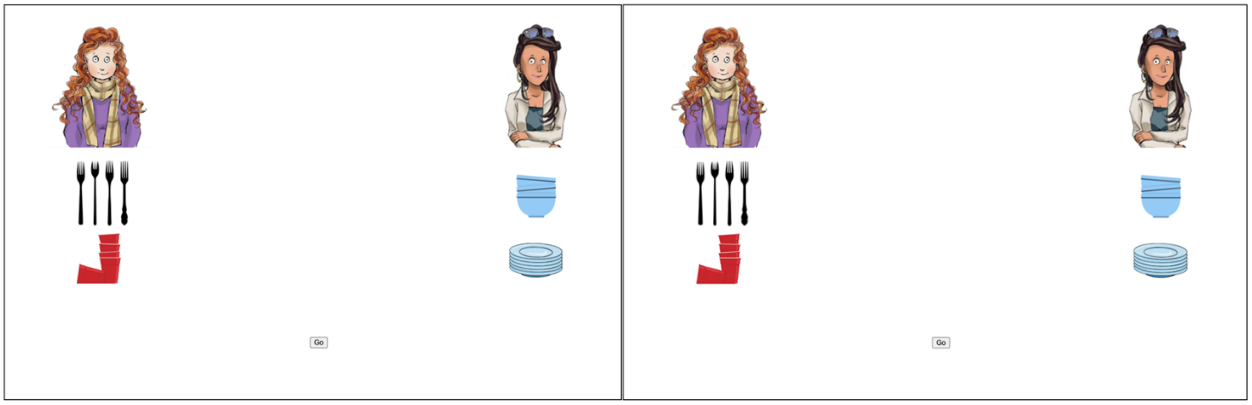

Figure 1. Example visual stimuli from pilot 1. *"Ana and Liz were cleaning up after the dinner party. Ana washed the towel with Liz and then she dried the plates."* Target is placed under subject character Ana (left) or under nonsubject character Liz (right).

Table 1. Example stimuli from pilot 1.

Item Type	Story
Practice	Liz had lunch at the cafeteria. She had a sandwich.
Exposure	Liz and Will joined an orchestra. Liz practiced with Will and then she accidentally broke her cello.
Critical	Ana and Liz were cleaning up after a dinner party. Ana washed the towel with Liz and then she dried the plates. ("plates" under Ana/Liz) Will and Matt were watching TV. Will had a snack with Matt and then he cleaned up the table. ("table" under Will/Matt)

Procedure

Participants accepted the HIT on Amazon Mechanical Turk and were directed to the PC-Ibex platform that scripted and run the experiment. The experiment began with a consent form, followed by a background questionnaire including the information mentioned above. The questionnaire was followed by the audio check to make sure participants could appropriately hear the spoken stimuli. If participants did not meet any of the inclusion criteria (age, age of learning English, English language proficiency or audio check), the experiment terminated, and participants were automatically dismissed from the study. After completion, participants received an Mturk code to be entered when submitted their work to receive payment.

Before the main task, participants were introduced to the four characters and told that they would be listening to short stories and answering questions. There was an example of the visual display that included two characters and two pairs of objects. They were told to only click on the object mentioned in the story and were required not to move too early. Each time they made a mistake there would be a feedback message popping out as a reminder of instruction. At the bottom-center of the screen there was a button labeled “go”. Participants needed to click the button to begin each trial, which made sure that at the beginning of each trial their mouse started from the same point on the screen. After clicking “go”, participants listened to the story, during which they could freely move their mouse anytime without warnings about moving too early, although this would be sent to the results file showing in which trial participants moved too early before the onset of the pronoun. However, they received a late warning “Responded too late.” if they responded 2 seconds after the offset of the critical word. During the experiment, the reaction times and mouse-trajectories from the starting point to the time participants made a click were recorded separately for each trial.

Analytical Approach

The measures of interest included overall accuracy, reaction times, and characteristics of mouse-trajectories. Each dependent measure is defined in detailed below. The accuracy served as an exclusion criterion before further analysis. For the online measures summarized in Table 2, the quantitative data were analyzed with a mixed-effects linear regression using SAS proc glimmix. The primary predictor is target location (under subject vs. under nonsubject) and was effects-coded as 1 and -1. Models included random intercepts for participant and item as well as random slopes. A secondary analysis included trial order as another predictor to explore whether it interacts with the main predictor, considering people would get used to the task and be increasingly efficient.

Accuracy. For all items including the critical stories that had ambiguous pronouns, there was only one correct answer indicated by the object picture. Therefore, accuracy was calculated as the percentage of correct selections out of the total number of items.

RT_off_critical. Reaction times in milliseconds were calculated as the time from the offset of the critical word to the time when participants made a click, in order to eliminate the potential confounds due to varied length of stimuli. This measure would indicate how long it took for participants to process the critical target, resolve any competition and make a final decision.

Mouse-trajectory parameters. For each participant and each trial, the real time *x*- and *y*-coordinates in pixels of mouse movements were recorded at a sampling rate of 60Hz. Recording started from the onset of each trial. Because participants might use different computers with differing screen sizes, all trajectories were rescaled into a standard coordinate space that retained the aspect ratio of most computer screens, shown in Figure 2 (Freeman et al., 2008; Freeman & Ambady, 2009; Freeman et al., 2010). According to most mouse-tracking studies in the literature, mouse trajectories can be used for analysis of spatial attraction, defined as how much the trajectories in one experimental condition travel closer to an unselected alternative (i.e., competitor of the target) relative to those in another experimental condition (Freeman & Ambady, 2009). There were four mouse-tracking measures of interest as the following. Each measure was analyzed separately.

(1) ***Initiation times (ms).*** Recorded as the moment the mouse was first moved, starting from the onset of the trial. Initiation times were used to explore individual variations in task strategies, which might signal whether participants generated expectations about upcoming information and act on them. For example, participants who actively anticipate the next-mentioned character were expected to start moving the mouse early in the trial even before they heard of the disambiguating word. In contrast, if participants did not move the mouse until the target word was revealed, it's possible that they were taking a strategy to just wait until the end. Also, initiation time might reveal differences between conditions, suggesting some items are more difficult than others that participants hesitated to make a selection.

Trajectories that travel significantly closer to the other objects rather than the target should display a spatial attraction to the distractors. Spatial attraction could be measured by the two most frequently used outcomes (e.g., Farmer et al., 2007) as the following:

(2) ***Maximum Absolute Deviation (MAD)***. Calculated as the largest perpendicular deviation between the actual trajectory and the straight trajectory (i.e., the shortest distance of movement towards the target, shown as the straight dashed line in Figure 2). The value is signed to indicate the direction of deviation. Positive if the MAD occurs above the straight trajectory and negative if it occurs below the straight trajectory. Greater value indicates stronger deviation from the target.

(3) ***Area Under the Curve (AUC).*** Calculated as the area between the actual trajectory and the straight trajectory. This measure was considered a complementary measure that is positively associated with MAD (Freeman et al., 2008). Positive if above the

straight trajectory and negative if below the straight trajectory. Greater value indicates stronger deviation from the target.

Finally, the streaming coordinates were also used for measuring complexity of mouse trajectories (e.g., Dale et al., 2008; Kieslich et al., 2019). Complexity is usually measured with the following metrics.

(4) ***X-flips_after_pronoun.*** X-flips were calculated as the number of directional reversals along the horizontal axis. This parameter could be used to measure the fluctuations in mouse movements to reveal any competition in processing among the alternative objects. In the current study, x-flips_after_pronoun were calculated by hand. For each trial, x-flips that occurred before the pronoun onset were subtracted from the total number of x-flips happening during the trial, resulting in our measure of x-flips_after_pronoun. We only included x-flips within the time window of interest to reduce the influence of cognitive processes irrelevant to the pronoun resolution process.

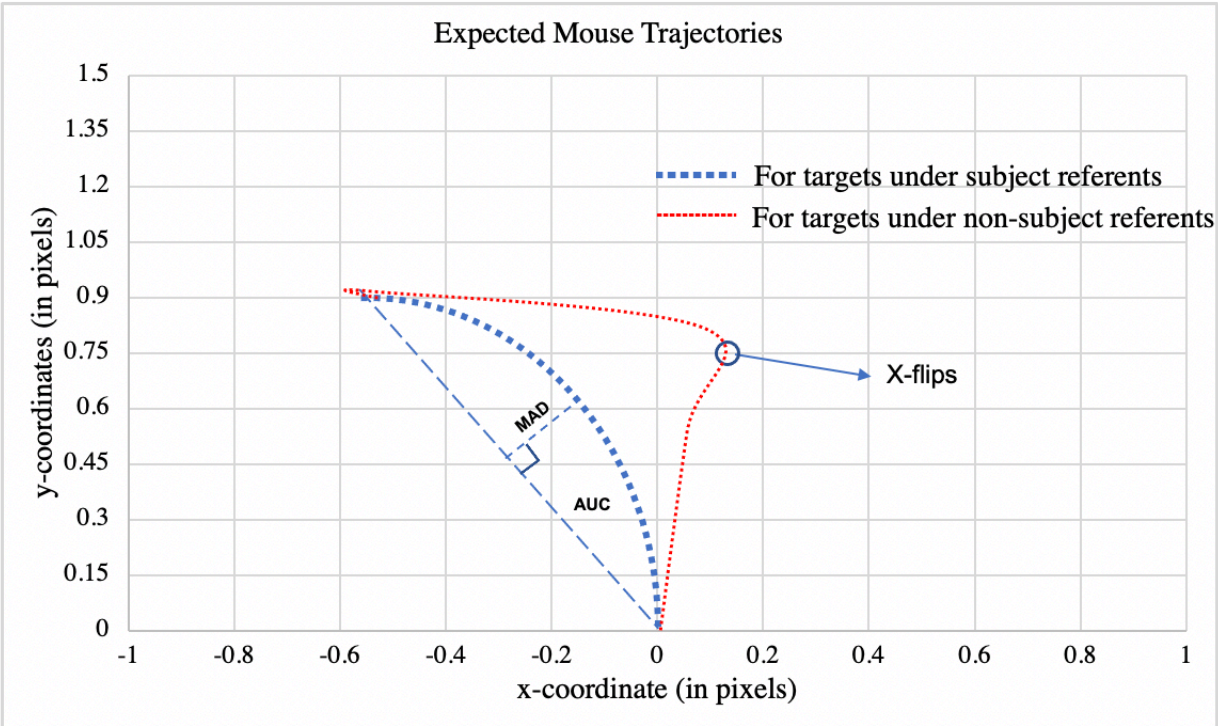

Figure 2. Predicted mouse trajectories for targets placed under expected (subject) vs. unexpected (nonsubject) referents. Graphs were adapted from Freeman & Ambady (2010).

Table 2. Summary of online measures including time and mouse-tracking parameters.

Time Parameters	**Definition**
RT_off_critical	From onset of the target word to the moment when mouse clicks on a selected object
Mouse-Trajectory Parameters	
Initiation Time	The moment when mouse movement initiates
Maximum Absolute Deviation (MAD)	Signed largest perpendicular deviation between the actual trajectory and the straight trajectory
Area Under the Curve (AUC)	Area between the actual trajectory and the straight trajectory
X-flips	Number of directional reversals along the horizontal axis

Data Inspection and Filtering

Before analysis, it was recommended that trials should be screened and filtered based on a series of criteria due to the large variability in mouse tracking data (Kieslich et al., 2019). Although mouse tracking has been shown as a sensitive measure to a variety of cognitive tasks, an accurate examination on the mapping between the cognitive processes and the mouse movements may require a dataset with less noisy trajectories (Calcagni et al., 2019). Based on the literature, we applied inspections to the raw dataset and the following exclusion criteria to filter out potentially biased trials.

Incorrect selections. Trials with clicks on any objects other than the targets were excluded to ensure that the trajectory being analyzed is related to the cognitive process being measured.

Early click. Responses before the onset of the pronoun were excluded.

RT_off_critical outliers. A reaction time after the offset of the critical word was calculated as an outlier if it was 2 standard deviations above the mean. Trials with a reaction time outlier were excluded in the RT analysis but not in other analyses.

We excluded 30 items for incorrect selections (6%). Of the correct items, 3 were excluded for early clicks (0.6%). This left 463 trials in the analyses. For the RT analyses only, we also excluded RT outliers (n=11, 2.3%).

Results

Accuracy. The overall accuracy was 93.95% across all participants. There was no difference between target-under-subject and target-under-nonsubject conditions (94.35% vs. 93.55%). Considering the simplicity of the task and that the target could be easily identified, the accuracy was not expected to be highly differential between conditions or to reveal any patterns of processing.

For online measures, after exclusions, only RTs, initiation times and x-flips_after_pronoun showed a slight numeric advantage for targets placed under subject characters, shown below as Figures 3, 4, and 5. However, none of these measures reached significance, summarized in Tables 3-7. In our secondary analyses that included order and its interaction with target location, only an order effect was found on AUC. The statistics were summarized in Tables A1-A5 in Appendix 1.

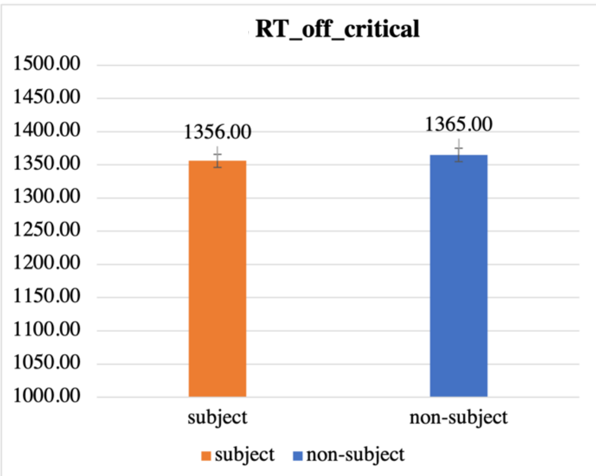

Figure 3. Mean RTs by condition for pilot 1 after exclusions.

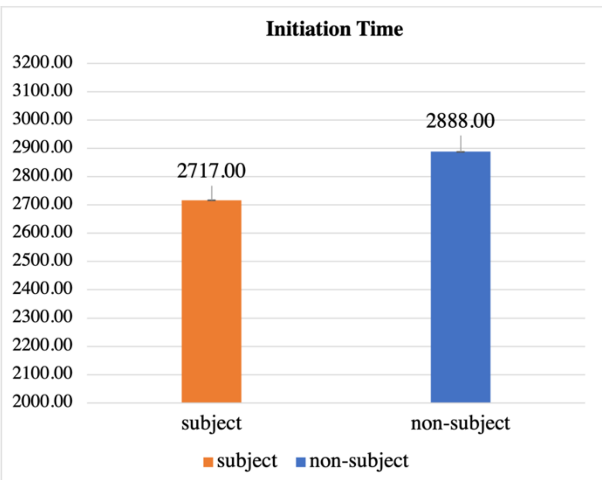

Figure 4. Mean initiation times by condition for pilot 1 after exclusions.

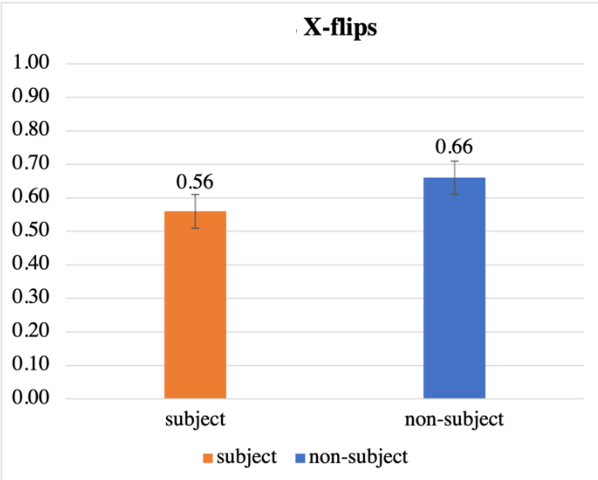

Figure 5. Mean x-flips_after_pronoun by condition for pilot 1 after exclusions.

Table 3. Inferential statistics for RT_off_critical from pilot 1.

Effect	Estimate (St. Error)	t Value	p
Intercept	1280.27 (49.36)	25.94	<.0001
Target location	-14.80 (32.37)	-0.46	0.65

Table 4. Inferential statistics for initiation times from pilot 1.

Effect	Estimate (St. Error)	t Value	p
Intercept	2838.01 (348.02)	8.15	<.0001
Target location	-41.3987 (92.32)	-0.45	0.66

Table 5. Inferential statistics for MAD from pilot 1.

Effect	Estimate (St. Error)	t Value	p
Intercept	40.10 (11.60)	1.77	3.46
Target location	4.26 (6.66)	-0.33	0.64

Table 6. Inferential statistics for AUC from pilot 1.

Effect	Estimate (St. Error)	t Value	p
Intercept	6722.94 (2139.74)	3.14	0.0038
Target location	-550.11 (1251.37)	-0.44	0.66

Table 7. Inferential statistics for x-flips_after_pronoun from pilot 1.

Effect	Estimate (St. Error)	t Value	p
Intercept	0.235 (0.05)	4.45	<.0001
Target location	-0.00255 (0.04)	-0.07	0.95

Discussion

Overall, none of the time or mouse-tracking measures revealed any reliable differences in processing of subject- and nonsubject-pronoun referential structures. It was expected that the online measures obtained from the mouse-tracking paradigm would have greater values, indicating stronger competition or difficulty during processing when the target was placed under the nonsubject character, as people were expected to look at or pay more attention to the subject character when they heard of the pronoun but not the critical word yet. However, pilot 1 failed to replicate the well-known subject bias found in studies using pronoun comprehension tasks.

PILOT 2

Pilot 2 was identical to pilot 1 except that the procedures were modified for the study to proceed in an experimental lab room. This pilot followed the procedures by Arnold et al. (2023), which is also an in-person study. This pilot served as a comparison of data collection approach and data quality to pilot 1, to explore whether different experimental conditions would influence the results.

Methods

Participants

21 undergraduate students from University of North Carolina at Chapel Hill participant pool participated in exchange for course credits. All participants reported as native speakers for learning English before age 6 and a native level of proficiency. No participants who completed the study were excluded for low accuracy. Therefore, all 21 participants were included in the final analysis.

Materials and Design

All stimuli and design were identical to pilot 1.

Procedure

All procedures were the same except for the following. First, the experiment was displayed on a lab computer with mouse sensitivity set to a low speed level of 3. As the study was conducted in-person, an experimenter was present throughout the whole experiment to introduce the study and read the instructions to the participants. During the experiment, the experimenter monitored if the participant 1) moved too early 2) made incorrect selections for more than 7 trials, 3) kept their hands on the mouse all the time, 4) got too many warnings about taking too long. For each scenario, the experimenter gave reminders accordingly to keep the participants follow the instructions, to reduce the effect of individual variations in ways of performing the task. The experimenter's monitoring of early movements was adopted from the metrics used in Arnold et al. (2023), in which they displayed an early warning whenever participants moved before the pronoun onset. This could potentially reduce influence of early mouse movements that are irrelevant to the pronoun resolution process.

Analytic Approach

All analyses were identical to pilot 1.

Data Inspection and Filtering

All exclusion criteria were identical to pilot 1. Here we excluded 6 items for incorrect selections (2%). No trials were excluded for early clicks. This left 330 trials in the analyses. For the RT analyses only, we excluded the RT outliers (n=13, 3.9%).

Results

Accuracy. The overall accuracy was 98.21% across all participants. There was no difference between target-under-subject and target-under-nonsubject conditions (98.81% vs. 97.62%).

For online measures, again, only RTs, initiation times and x-flips_after_pronoun showed a slight numeric advantage in processing of targets placed under subject characters, shown below as Figures 6, 7, and 8. However, none of these measures reached significance, summarized in Tables 9-12. Again, in secondary analyses with order being added as a predictor, only an order effect was found in analyses of AUC. Statistics were summarized in Tables A6-A10 in Appendix 1.

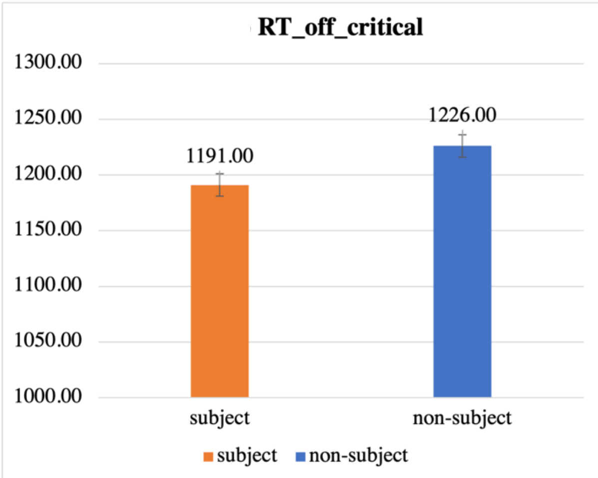

Figure 6. Mean RTs by condition for pilot 2 after exclusions.

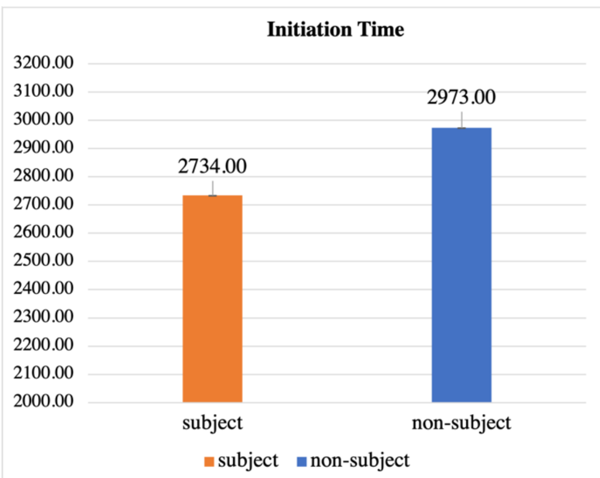

Figure 7. Mean initiation times by condition for pilot 2 after exclusions.

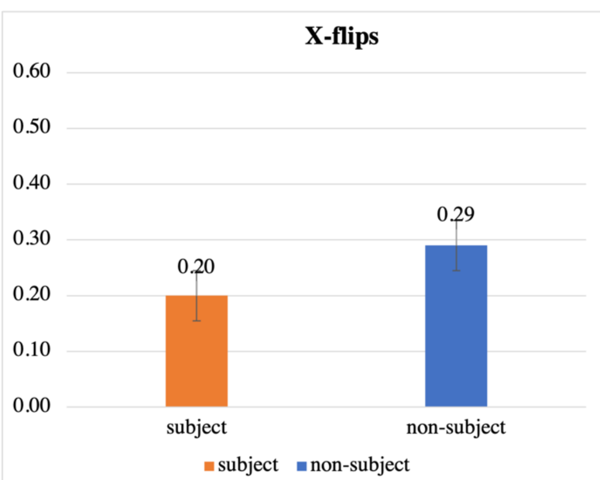

Figure 8. Mean x-flips_after_pronoun by condition for pilot 2 after exclusions.

Table 8. Inferential statistics for RT_off_critical from pilot 2.

Effect	Estimate (St. Error)	t Value	p
Intercept	1216.51 (54.51)	22.32	<.0001
Target location	-17.29 (33.99)	-0.51	0.62

Table 9. Inferential statistics for initiation time from pilot 2.

Effect	Estimate (St. Error)	t Value	p
Intercept	2787.55 (344.93)	8.08	<.0001
Target location	-93.30 (126.82)	-0.74	0.47

Table 10. Inferential statistics for MAD from pilot 2.

Effect	Estimate (St. Error)	t Value	p
Intercept	30.71 (7.98)	3.85	0.0008
Target location	6.71 (6.95)	0.97	0.36

Table 11. Inferential statistics for AUC from pilot 2.

Effect	Estimate (St. Error)	t Value	p
Intercept	6722.94 (2193.74)	3.14	0.0038
Target location	-550.11 (1251.37)	-0.44	0.66

Table 12. Inferential statistics for x-flips_after_pronoun from pilot 2.

Effect	Estimate (St. Error)	t Value	p
Intercept	0.24 (0.06)	3.97	0.0007
Target location	-0.04 (0.04)	-1	0.33

Discussion

Similar to the results in pilot 1, pilot 2 did not find any significant differences in processing between conditions, even though this pilot had a more strictly-monitoring experimental environment and participants did seem to be more attentive during the experiment, suggested by the lower number of exclusions and the accuracy rate, which was almost 6% higher than that of the Mturk participants. However, this pilot had only 21 participants and might have been under powered. Overall, as this paradigm did not seem to be working in this investigation, the pilots were not followed up to test referential adaptation to the subject- versus nonsubject-pronoun referential structures. Several possibilities for why this mouse-tracking paradigm failed to replicate the subject bias are proposed in the general discussion.

MAIN EXPERIMENT

The main experiment aimed to follow up the findings by Arnold et al. (2023), using mouse-tracking to replicate the adaptation effects observed in offline behavioral patterns by specifically testing the effect of exposure on processing *they* pronouns. Nonbinary singular they for its low-frequency and relatively new status in pronoun comprehension, involves a processing cost relatively other third-person singular (e.g., *he/she*) and plural *they* pronouns. This study followed Arnold et al.'s (2023) study design and extended it to add an exposure manipulation.

Methods

Participants

A power analysis was conducted using PROC Power in SAS. Based on the effect size for RTs and x-flips in Arnold et al. (2023), a target sample of 96 participants was determined as necessary to detect the predicted interaction at $p < .05$ 80% of the time. 124 Amazon Mechanical Turk workers participated in exchange for $2.5. The Mturk HIT was only available to native English speakers from the US. The participants were required to have 1000 previous HITs with at least 95% acceptance rate. All participants reported as native speakers for learning English before age 6 and a native level of proficiency. Out of all 124 participants who completed the study, 27 were excluded for low accuracy, leaving 97 participants in the final analysis. The excluded participants achieved lower than 85% accuracy across all items and were considered inattentive.

Materials and Design

Identical to pilot 1, the experiment was implemented on web via the PC-Ibex platform, beginning with background demographic questions and the then main story task. Demographic questions asked about age, sex, gender, ethnicity/race, language experience, language proficiency and report of language disorders. To ensure the spoken stimuli would be received appropriately, participants were given an audio test ("The number is seven.") and asked to click on one of the numbers from 1 to 7. The experiment would terminate if participants did not choose the mentioned number.

This study used linguistic stimuli adapted from Arnold et al. (2023). For the main task, the setup of the experiment is illustrated in Figure 9. The basic paradigm introduces 3 characters and their pronouns (Alex-*they*, Liz-*she*, Will-*he*) and a series of short stories. Participants were explicitly told that the character Alex use *they* pronouns, to make participants aware of the pronoun status. Each story consists of two context sentences introducing 2 characters, followed by another sentence with a pronoun referring to either the singular they character or the plural characters, as shown in Table 13. For each story, three pictures of characters were placed at bottom-left, bottom-middle, and bottom-right respectively, two pictures of individual characters and one picture of two characters together. A pair of objects was placed under each of the character pictures respectively. The visual objects could be potentially fit into the sentence but only one target object was mentioned in the spoken story. The target object placed under one of the character pictures to indicate the interpretation of the given pronoun.

There were 20 critical stories that always had Alex and one of the other characters, such that the target object was either under Alex to indicate a singular interpretation of *they* or under the two characters together to indicate a plural interpretation of *they*. The critical stories always used source-goal verbs in the second sentence where Alex was always the goal character in the transfer event. This makes the singular interpretation more natural because in general people are likely to expect the goal character to be mentioned next. The 24 exposure stories were manipulated so that the target object was either always under Alex (singular exposure) or always under the plural characters (plural exposure). For the 20 fillers, the stories only had Liz or Will, varied in the number of characters mentioned that either only one character was mentioned or both, and varied in the location of targets that were either under the subject or the nonsubject antecedent. However, regardless of story structure, the visual setting remained consistent across all the trials that each had three character pictures and three object pairs.

Overall, the experiment had a 2 by 2 design with a within-subject manipulation of target location (under singular vs. under plural pictures) and a between-subject manipulation of exposure type (singular *they* exposure vs. plural *they* exposure). This design led to a total of 4 lists. Shown in Table 13, the exposure items included two story types. For singular exposure, half of the stories had a context with Alex alone and half had two characters, including Alex with one of the other characters. For plural exposure, half of the stories had a story context with Liz and Will together and half had Alex with one of the other characters. In addition, for all two-character stories, Alex was always the second-mentioned. As suggested by Arnold et al. (2023)'s findings, participants experienced strongest difficulty of processing singular they when Alex was the second-mentioned, as the subject bias would elicit further competition with the singular interpretation. Again, the number of times the targets objects and target characters were placed left or right was counterbalanced.

Figure 9. Example visual stimuli from the main experiment. *"Liz and Alex were at a restaurant. Liz handed the menu to Alex. Then they ordered some wine."* Target "wine" is placed under the singular character Alex (left) or under the plural characters (right).

Table 13. Example stimuli from main experiment.

Item Type	Story
Practice	Alex and Liz went ice skating. Alex had to go back home. They forgot their skates.
Exposure:	
Singular They	**Alex-Only:** Alex went to the store. They looked through all the aisles. Then they bought some milk. (Target under Alex) **Alex-Second:** Will and Alex were having lunch at the park. Will passed a fancy box to Alex. Then they ate a piece of sushi. (Target under Alex)
Plural They	**Liz-and-Will:** Will and Liz went to the store. They looked through all the aisles. Then they bought some milk. (Target under Plural Picture) **Alex-Second:** Will and Alex were having lunch at the park. Will passed a fancy box to Alex. Then they ate a piece of sushi. (Target under Plural Picture)
Critical:	
Singular Interpretation	Will and Alex were at marching band practice. Will gave some water to Alex. Then they put down the batons. (Target under Alex)
Plural Interpretation	Will and Alex were at marching band practice. Will gave some water to Alex. Then they put down the batons. (Target under Plural Picture)

Procedure

The procedures of this experiment were identical to pilot 1, which was also conducted on Mturk, except for the following. First, following the background questionnaire, participants were introduced to three characters and were explicitly told about their pronoun use (Alex-*they*, Liz-*she*, Will-*he*). Second, the visual context displayed three character pictures with three pairs of objects. Participants needed to click on the target object out of a total of six object pictures. More importantly, this study added an early warning message at the end of trial, as Arnold et al. (2023) did, whenever they moved before the pronoun onset. This message reminded participants to try to stay still until they really knew where to move, and thus could efficiently reduce the rate of free movements that were irrelevant to the cognitive processes of interest.

Analytic Approach

In this experiment, we measured the accuracy of responses, RT_off_critical, and mouse-tracking measures, including MAD, AUC and x-flips_after_pronoun. For RT_off_critical, outliers were calculated as 4 standard deviations above the mean. We set a lower threshold for excluding outliers than we did in pilots 1 and 2 because processing singular vs. plural they references was more difficult than processing subject vs. nonsubject references. Additionally, the RT outliers were only excluded for RT analysis but not for the other measures. The quantitative data were analyzed with a mixed-effects linear egression using SAS proc glimmix. The primary predictors were target location (under singular vs. under plural) and exposure type (singular exposure vs. plural exposure). The predictors were effects-coded as 1 and -1. Models included random intercepts for participant and item as well as random slopes. A secondary analysis included trial order as another predictor as well as all interactions between order and the predictors.

Data Inspection and Filtering

Identical to the pilots, trials with incorrect selections, early clicks (response before critical word) were removed from analysis. Further, only for RT_off_critical, trials with RT outliers (4 standard deviations above the mean) were removed from analysis. We excluded 82 trials for incorrect selections (4.23%). Of all the correct items, 20 were excluded for early clicks (1.03%). This left 1858 trials in the analyses. For the RT analyses only, we excluded the RT outliers (n=17, 0.88%).

Results

Accuracy. The overall accuracy of the experiment was 95.77%, with no significant difference between conditions (plural_exposure-target_under_plural: 95.83% vs. plural_exposure-target_under_singular: 94.38%; singular_exposure-target_under_plural: 96.94% vs. singular_exposure-target_under_singular: 95.92%).

RT_off_critical. The numeric data displayed an expected cross-over interaction between exposure types and target locations. Specifically, when participants received plural exposure they took longer to process and respond in trials with targets placed under the singular character Alex. This pattern was reversed in the other condition where people received singular exposure. Participants were faster processing targets placed under singular characters. In support of the numerical patterns, we observed a marginal interaction between exposure and target location, p = .079, shown in Table 14. It is possible that there was an exposure effect leading to this cross-over interaction, but the effect size was too small for observing significance. Additionally, for singular exposure, the RTs for singular interpretations were no difference from RTs for the singular in plural exposure conditions. However, the exposure type was a between-subject manipulation that could not directly compare the two groups. Therefore, this lack of exposure effect between groups might result from the large variations in individual differences.

In a secondary analysis for RT_off_critical, order, and its two-way interactions with each predictor, as well as its three-way interactions with the two predictors were added to the model. Again, no significant effect or interaction was found. Statistics were summarized in Table A11 in Appendix 1.

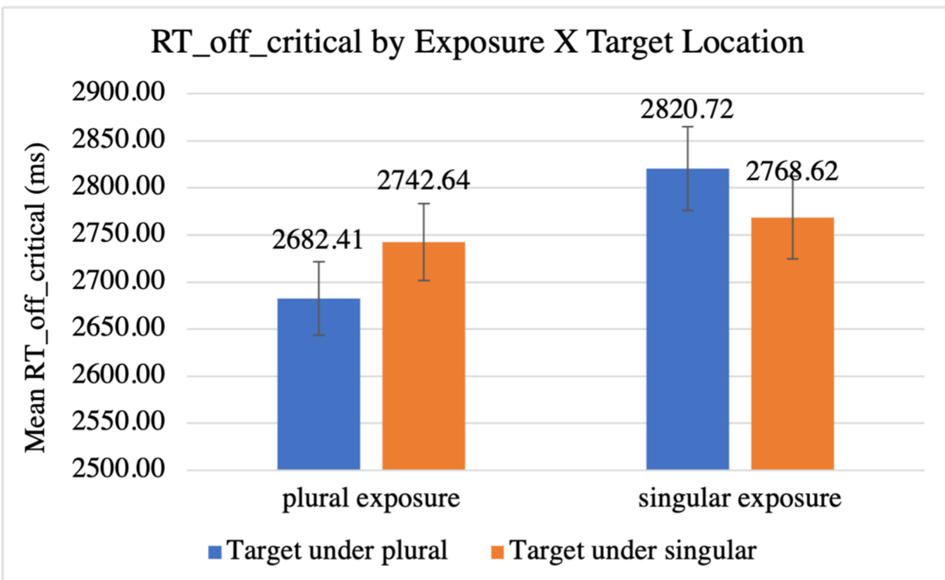

Figure 10. The means of RT_off_critical by exposure type and target location.

Table 14. Inferential Statistics for RT_off_critical (basic model without order).

Effect	Estimate (St. Error)	t Value	p
Intercept	2766.81 (86.69)	31.92	<.0001
Exposure	40.19 (59.11)	0.68	0.50
Target Location	6.22 (15.78)	0.39	0.69
Exposure X Target Location	-29.57 (17.54)	-1.87	0.079

AUC. The means of AUC showed an expected direction of numeric pattern in the singular exposure condition. The AUC was much lower for targets under singular characters than for targets under plural characters, possibly indicating a weaker competition for singular interpretations when given singular exposure. However, none of the effects was statistically significant, as reported in Table 15. Additionally, as shown in Table 16, the order was significant in the full model, showing possibly 1) competition was decreasing over time across the experiment and 2) participants were adapting to the task and became more and more efficient over time.

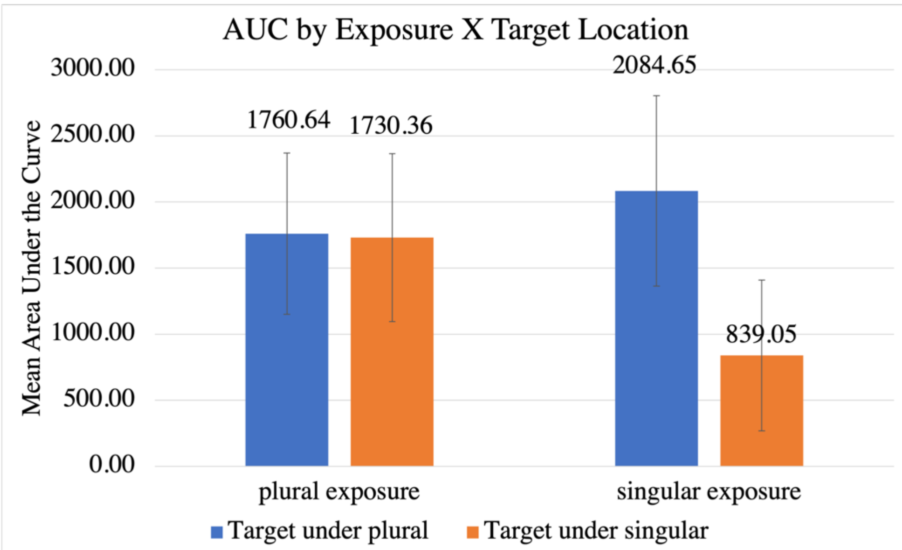

Figure 11. Mean area under the curve by conditions (Exposure X Target Location).

Table 15. Inferential statistics for AUC (basic model without order).

Effect	Estimate (St. Error)	t Value	p
Intercept	1644.32 (596.40)	2.76	0.01
Exposure	-151.83 (633.72)	-0.24	0.81
Target Location	-287.19 (307.64)	-0.93	0.35
Exposure X Target Location	-329.88 (307.64)	-1.07	0.29

Table 16. Inferential statistics for AUC (secondary model with order)

Effect	Estimate (St. Error)	t Value	p
Intercept	-378.23 (981.80)	-0.39	0.7033
Exposure	-273.52 (666.47)	-0.41	0.6836
Target Location	-79.122 (716.29)	-0.11	0.9121
Exposure X Target Location	-366.08 (325.97)	-1.12	0.2633
Order	58.3447 (22.419)	2.6	0.0169
Exposure X Order	-2.4012 (17.573)	-0.14	0.8913
Target Location X Order	-378.23 (20.27)	-0.39	0.7033
Exposure X Target Location X Order	-273.52 (666.47)	-0.41	0.6836

MAD. Similar to RT_off_critical, the numeric pattern showed a cross-over interaction between exposure and target location, such that participants exposed to plural they experienced stronger competition for singular interpretations. In contrast, for participants receiving singular exposure, they instead experienced stronger competition for plural interpretations. This interaction was marginally significant in the basic model, p = .07, shown in Table 17. More importantly, follow-up simple effects test (Table 19) showed a significant difference between target location conditions in singular exposure condition. Therefore, the singular exposure elicited a significant advantage for processing targets placed under the singular character. Additionally, there was a main effect of order in secondary analyses (Table 18).

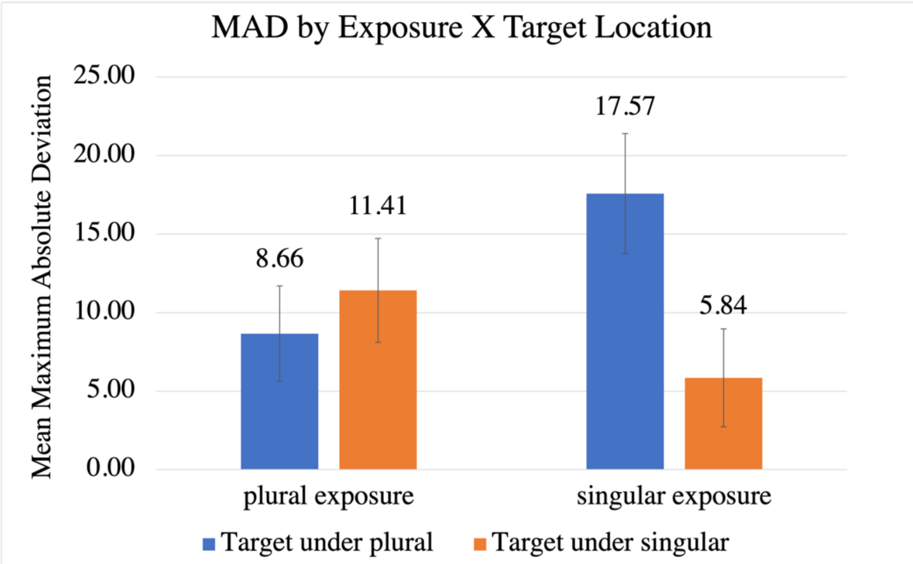

Figure 11. Mean maximum absolute deviation by conditions (Exposure X Target Location).

Table 17. Inferential statistics for MAD (basic model without order).

Effect	Estimate (St. Error)	t Value	p
Intercept	11.81 (3.76)	3.14	0.003
Exposure	1.11 (3.58)	0.31	0.76
Target Location	-1.77 (1.61)	-1.09	0.28
Exposure X Target Location	-3.32 (1.71)	-1.94	0.07

Table 18. Inferential statistics for MAD (secondary model with order)

Effect	Estimate (St. Error)	t Value	p
Intercept	-0.58 (5.69)	-0.37	0.7127
Exposure	-4.37 (5.71)	-1	0.3292
Target Location	-1.32 (3.61)	-0.83	0.4165
Exposure X Target Location	0.39 (3.79)	-0.16	0.8759
Order	0.33 (0.13)	2.93	0.0084
Exposure X Order	0.15 (0.13)	1.35	0.1933
Target Location X Order	-0.01 (0.09)	0.24	0.8097
Exposure X Target Location X Order	-0.101 (0.10)	-0.93	0.3651

Table 19. Estimates for MAD in basic model (simple effects of interaction)

Effect	Estimate (St. Error)	t Value	p
Singular Exposure Effect on Target Location	-10.18 (4.66)	-2.18	0.035
Plural Exposure Effect on Target Location	3.11 (4.74)	0.66	0.52

X-flips_after_pronoun. Finally, the pattern of x-flips also showed the similar pattern but now the effect showed up in the plural exposure condition, that the plural exposure elicited an advantage for plural against singular *they*, indicated by the marginal significant interaction of exposure and target location, p = .069. The interaction was further probed in the follow-up simple effects test, showing a marginal plural exposure effect on interpretations of singular vs. plural *they*, p = .067. The statistics were summarized in Tables 20-22.

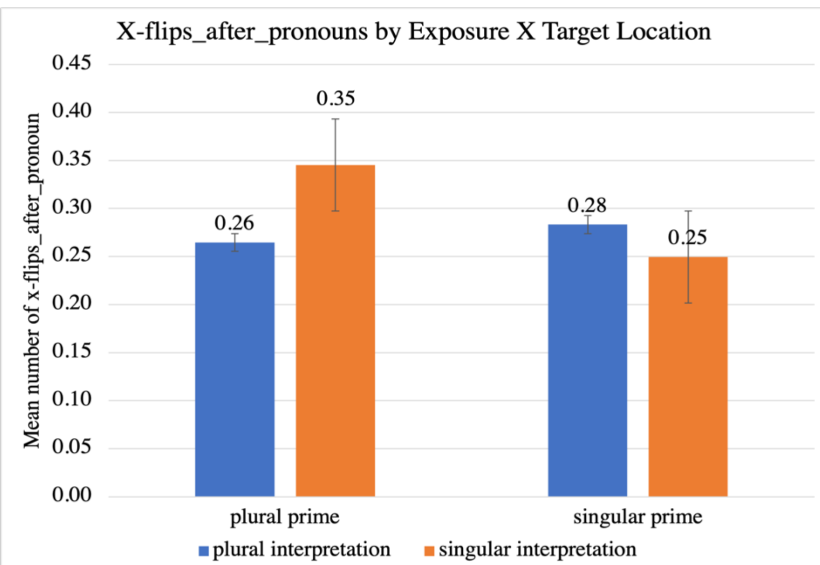

Figure 11. Mean x-flips_after_pronoun by conditions (Exposure X Target Location).

Table 20. Inferential statistics for x-flips (basic model without order).

Effect	Estimate (St. Error)	t Value	p
Intercept	0.30 (0.03)	9.13	<.0001
Exposure	-0.02 (0.03)	-0.71	0.4826
Target Location	0.01 (0.02)	0.88	0.3901
Exposure X Target Location	-0.03 (0.01)	-1.82	0.0692

Table 21. Inferential statistics for x-flips (secondary model with order)

Effect	Estimate (St. Error)	t Value	p
Intercept	0.29 (0.05)	5.55	<.0001
Exposure	0.07 (0.04)	1.58	0.12
Target Location	0.06 (0.03)	1.69	0.11
Exposure X Target Location	-0.00 (0.03)	-0.03	0.97
Order	0.00 (0.00)	0.08	0.94
Exposure X Order	-0.00 (0.00)	-2.78	0.01
Target Location X Order	-0.00 (0.00)	-1.45	0.17
Exposure X Target Location X Order	-0.00 (0.00)	-0.83	0.41

Table 22. Estimates for x-flips in basic model (simple effects of interaction)

Effect	Estimate (St. Error)	t Value	p
Singular Exposure Effect on Target Location	-0.02 (0.04)	-0.59	0.5549
Plural Exposure Effect on Target Location	0.08 (0.04)	1.87	0.0668

Discussion

The results together showed a marginal interaction between exposure and target location for RTs, MAD and x-flips_after_pronoun, which are measures of competition during processing in a mouse-tracking paradigm. This cross-over interaction suggests that the different interpretations of *they* were competing with each other during processing, and a specific exposure pattern led to stronger competition in processing of the other non-exposure interpretation of *they.* Additionally, a consistent effect of order was found, such that in general people tended to get more efficient doing the task across the experiment. Overall, there was a marginal effect of exposure on online processing during adaptation to referential structures of t*hey*. However, issues arise for using mouse-tracking in this investigation. This paradigm might not be sensitive enough to capture the robust effect found in behavioral results. Very few trials show any anticipation of participants. The majority of participants moved after the critical word offset, suggesting that most participants adopted strategy of waiting until the answer to be revealed without actively anticipating for re-mentioned referents.

General Discussion

The current study explored the use of mouse-tracking in the investigation of referential adaptation. Following up Johnson & Arnold (2022) and Arnold et al. (2023), this study focused on 1) does mouse-tracking reveal any referential processing of subject- and nonsubject- pronoun structures? and 2) do people adapt to singular *they*? To answer these questions, we tested 1) the differences in online processing of subject- vs. nonsubject-pronoun structures by using Arnold et al. (2023)'s mouse-tracking paradigm, and 2) online adaptation to nonbinary singular *they* by integrating mouse-tracking with Johnson and Arnold's referential adaptation paradigm. Pilots 1 and 2 testing the referential processing of subject vs. nonsubject references failed to replicate the robust subject bias. It remains unknown whether the current paradigm is sensitive to the processing of all kinds of referential structures or only those eliciting strong ambiguity and competition. The main experiment found a marginal effect of exposure that depends on what interpretations people need to make. The results suggest that exposure to one of the *they* interpretations tends to increase competition against the other interpretation.

The first question of the current study asked whether our mouse-tracking paradigm is sensitive to referential processing in general by testing how it reveals differences in processing subject vs. nonsubject interpretations of pronouns. Both pilot studies failed to reveal any reliable sensitivity to the first-mention bias that has been quite robust in both offline behavioral responses and eye-tracking patterns (e.g., Arnold et al., 2010). On the one hand, it is possible that the resolution of the ambiguity between subject and non-subject antecedents is rapid and requires very low level of cognitive demand, so the competition between the two is too small to be captured by hand movements. This was demonstrated by Arnold et al. (2000), which is an eye-tracking study that tested how people integrated subject bias during pronoun resolution. In their study, they presented to participants pictures of two cartoon characters Mickey and Donald and stories like “Donald is bringing some mail to Mickey. While a violent storm is beginning. He’s carrying an umbrella, and it looks like they’re both going to need it.” The story had a critical item “umbrella” to disambiguate the pronoun that could potentially refer to either character. The critical item was either in Donald’s hand or in Mickey’s hand. What they found is that the subject bias leads to a rapid divergence in participants’ eye fixations, in that they looked more to the subject character as soon as 200ms after the pronoun onset. However, as soon as the disambiguating information was given, this bias was quickly revised and participants fixated on the target character. Therefore, it is likely that the online processing disruption caused by subject bias was too brief to be captured by our paradigm.

Therefore, this raises questions about how sensitive our mouse-tracking paradigm is compared to the eye-tracking paradigm and how quickly and appropriately mouse movements could be mapped onto the moment-by-moment cognitive processes. Certainly, it is possible to capture the effect by employing additional metrics to ensure the validity of data and increase the sensitivity of mouse-tracking of capturing real-time dynamics of hand movements. Several factors might be relevant.

Participants adopted task strategies. For a trajectory to accurately capture the on-going cognitive processes, the task relies on participants’ attentiveness. Attentive participants may be more likely to anticipate and start moving as they engage in an ongoing resolution process. In contrast, people may wait until the critical word is revealed and resolved by the visual scene. In this case, the movements will not reveal any ongoing processes and thus reduce the difference between conditions. However, it's difficult to distinguish these irrelevant trajectories.

Participants not following instructions. Although we have given explicit instructions (e.g., "*stay still until ready to move"*; "*try to respond as quickly as possible"*), the task is relatively unnatural – participants need to move early while not too early. Even with this restriction implemented, we still found 115 out of 463 trials with substantial early movements along the horizontal axis that might have biased our results. This influence was mitigated in pilot 2 by running mouse-tracking in lab with an experimenter monitoring the whole experiment. There were only 14 out of 330 trials with substantial early movements, indicating that in general participants were holding the mouse still before pronoun set. Comparing the exclusion rates for pilots 1 and 2, the ratios of exclusions in in-lab study were relatively lower than those in the online study, shown in Table 5 in pilot 1 and Table 15 in pilot 2 (Incorrect Selections: 33 vs. 6; Early Clicks: 6 vs. 0).

Overall, it's challenging to use mouse-tracking to capture very early processes. We have implemented several metrics, including "late warnings" to encourage faster responses and the strict data exclusion criteria. However, our results demonstrated that it would not be promising to use mouse-tracking to investigate referential processing of subject- and nonsubject references. Therefore, we did not follow up on the pilots but focused on adaptation to the singular- and plural-*they* references.

Then the next question asked in this study was whether mouse-tracking can reveal any referential adaptation effects to nonbinary singular *they*, which is a low-frequency pronoun use compared to the commonly used plural *they*. As found in Arnold et al. (2023), mouse-tracking is sensitive to the differential processing of the singular vs. plural *they* that their results of RTs and x-flips demonstrated significant difficulty processing singular *they*. Based upon this finding, the current study integrated mouse-tracking with Johnson and Arnold (2022)'s referential adaptation paradigm, by exposing participants to either singular *they* or plural *they*. The results of RTs, MAD and x-flips demonstrated a cross-over interaction between exposure and different interpretations of the pronouns. These results are marginal, but showed that to some extent people experienced stronger competition and difficulty for the non-exposure pronoun interpretation pattern.

This adaptation may be driven by two types of probabilities in the local context. First of all, the expectancy hypothesis by Arnold (1998; 2010) pointed out that a referent that's more likely to be re-mentioned next has a higher expectancy and thus is more accessible in the discourse. In the experiment, participants may keep track of how frequently a referent was going to be mentioned next, either the singular character Alex or the plural characters. The most frequent exposure structure leads to changes in participants' expectations about who's being mentioned next. Thus, singular exposure leads to higher expectation for Alex being re-mentioned and plural exposure leads to higher expectation for plural characters. On the other hand, people may keep track of how frequently a *they* pronoun refers to singular or plural referents. In this case, people are learning a post-hoc bias about who being referred to given a *they* pronoun. The current study is not able to distinguish which bias people are learning. Future research may further investigate this nature of referential adaptation.

There are also open questions about the mechanisms underlying referential adaptation. Some possible mechanisms may be proposed based on accounts for syntactic priming and adaptation, which includes the activation-based theories and implicit learning accounts (e.g., Pickering & Branigan, 1998; Bock & Griffin, 2000; Chang et al., 2012), although we cannot tell which one is involved based on current evidence. The activation-based theories explain priming and adaptation as a result of increasing residual activation of a previously encountered structure. In the case of referential adaptation, one possibility is that exposure leads to increasing activation of one interpretation of *they* over the other, such that exposure to singular *they* increases the activation level of a singular interpretation in competition with a plural interpretation of *they*. These interpretations may be represented as two different lexical meanings of *they*, or as different procedural knowledge of how *they* is resolved. Proposed by activation-based theories, this activation is short-lived and decay rapidly. However, for previous behavioral studies on referential adaptation, the effect of exposure was found to be long-lasting up to several days (e.g., Contemori et al., 2021). The longer-lasting effects of exposure may be better explained by the implicit learning accounts, which emphasize the long-term changes in one's mental representations of the exposure structures. These accounts are error-based, suggesting that infrequent structure elicits a prediction error. In order to reduce this error people can quickly adapt to the structure updating expectation about its distribution probability (e.g., Jaeger & Snider, 2013). This type of learning may also apply to the case of referential adaptation in this experiment, because the singular *they* is a relatively low-frequency pronoun use compared to the plural *they*. However, we cannot tell whether people are experiencing a prediction error and if so, what specifically this prediction error would be.

Overall, this study shows that exposure does not only affect offline pronoun comprehension but also influences online processing of referential structures. Aligning with the behavioral tendency to follow the exposure referential structure, people were faster processing those exposure structures while experiencing increasing competition with the non-exposure structure. This study also demonstrates the challenges to use the mouse-tracking to detect early processes. What we are seeing might be a long-lasting effect of exposure because the majority of participants tended to respond after the critical word offset.

APPENDIX 1: ADDITIONAL TABLES

Table A1. Inferential statistics for RT_off_critical by Target Location X Order in pilot 1.

Effect	Estimate (St. Error)	t Value	p
Intercept	1256.97 (80.69)	15.58	<.0001
Target Location	72.79 (73.78)	0.99	0.333
Order	0.92 (2.66)	0.35	0.7318
Target Location X Order	-3.43 (2.60)	-1.32	0.1997

Table A2. Inferential statistics for MAD by Target Location X Order in pilot 1.

Effect	Estimate (St. Error)	t Value	p
Intercept	13.26 (14.63)	0.91	0.3743
Target Location	10.99 (14.61)	0.75	0.4586
Order	1.06 (0.63)	1.68	0.1017
Target Location X Order	-0.26 (0.52)	-0.5	0.6202

Table A3. Inferential statistics for AUC by Target Location X Order in pilot 1.

Effect	Estimate (St. Error)	t Value	p
Intercept	4805.65 (3250.85)	1.48	0.143
Target Location	4732.4 (2821.97)	1.68	0.0943
Order	76.59 (101.96)	0.75	0.4545
Target Location X Order	-210.79 (-210.79)	-2.02	0.044

Table A4. Inferential statistics for X-Flips by Target Location X Order in pilot 1.

Effect	Estimate (St. Error)	t Value	p
Intercept	0.22 (0.10)	2.32	0.0259
Target Location	0.08 (0.09)	0.93	0.3582
Order	0.00 (0.00)	0.17	0.867
Target Location X Order	-0.00 (0.00)	-1.07	0.2928

Table A5. Inferential statistics for Initiation Times by Target Location X Order in pilot 1.

Effect	Estimate (St. Error)	t Value	p
Intercept	3191.72 (390.5)	8.17	<.0001
Target Location	-342.07 (201.55)	-1.7	0.0904
Order	-13.98 (7.34)	-1.9	0.0629
Target Location X Order	11.84 (7.18)	1.65	0.0999

Table A6. Inferential statistics for RT_off_critical by Target Location X Order in pilot 2.

Effect	Estimate (St. Error)	t Value	p
Intercept	1312.89 (71.54)	18.35	<.0001
Target Location	-83.81 (62.45)	-1.34	0.1911
Order	-4.46 (2.83)	-1.57	0.1253
Target Location X Order	3.04 (2.56)	1.19	0.2464

Table A7. Inferential statistics for MAD by Target Location X Order in pilot 2.

Effect	Estimate (St. Error)	t Value	p
Intercept	36.18 (12.41)	2.92	0.0071
Target Location	18.40 (11.76)	1.56	0.1314
Order	-0.25 (0.51)	-0.49	0.6279
Target Location X Order	-0.54 (0.53)	-1.02	0.3167

Table A8. Inferential statistics for AUC by Target Location X Order in pilot 2.

Effect	Estimate (St. Error)	t Value	p
Intercept	8818.42 (2915.18)	3.03	0.0051
Target Location	5925.19 (2810.01)	2.11	0.0946
Order	-97.39 (113.11)	-0.86	0.3974
Target Location X Order	-242.48 (113.35)	-2.14	0.0423

Table A9. Inferential statistics for X-Flips by Target Location X Order in pilot 2.

Effect	Estimate (St. Error)	t Value	p
Intercept	0.24 (0.09)	2.5	0.0172
Target Location	0.02 (0.08)	0.26	0.7971
Order	0.00 (0.00)	0.05	0.9617
Target Location X Order	-0.00 (0.00)	-0.84	0.4092

Table A10. Inferential statistics for Initiation Times by Target Location X Order in pilot 2.

Effect	Estimate (St. Error)	t Value	p
Intercept	2812.17 (407.2)	6.91	<.0001
Target Location	232.22 (250.8)	0.93	0.363
Order	-1.16 (10.25)	-0.11	0.9115
Target Location X Order	-15.43 (10.27)	-1.5	0.145

Table A11. Inferential statistics for RT_off_critical (secondary model with order) in main experiment.

Effect	Estimate (St. Error)	t Value	p
Intercept	2896.58 (163.17)	17.75	<.0001
Exposure	30.02 (73.85)	0.41	0.69
Target Location	12.16 (35.20)	0.35	0.73
Exposure X Target Location	-35.24 (3.81)	-1.00	0.32
Order	-3.44 (1.53)	-0.9	0.38
Exposure X Order	0.25 (0.86)	0.17	0.87
Target Location X Order	-0.23 (0.86)	-0.27	0.78
Exposure X Target Location X Order	0.08 (0.86)	0.09	0.93

APPENDIX 2: EXPERIMENT STIMULI

Table B1. Stimuli for pilots 1 & 2

Item	Context Sentence	Pronoun Sentence
Critical	Ana and Liz were cleaning up after a dinner party.	Ana washed the towel with Liz and then she dried the plates.
Critical	Will and Matt were watching TV.	Will had a snack with Matt and then he cleaned up the table.
Critical	Liz and Ana were grocery shopping.	Liz read the grocery list with Ana and then she grabbed a bag of chips.
Critical	Will and Matt went to the airport.	Will went through security with Matt and then he got some coffee.
Critical	Matt and Will went hiking.	Matt opened up a backpack with Will and then he packed some snacks.
Critical	Liz and Ana went to a bakery.	Liz counted the money with Ana and then she bought a croissant.
Critical	Matt and Will were getting ready to go to the beach.	Matt packed beach supplies with Will and then he forgot his sunscreen.
Critical	Liz and Ana were at a restaurant.	Liz read the menu with Ana and then she ordered some wine.
Critical	Matt and Will went to a restaurant.	Matt chatted with Will and then he ate pasta.
Critical	Matt and Will were studying together.	Matt wrote down some notes with Will and then he looked through the textbook.
Critical	Will and Matt went to a farm.	Will saw the cows with Matt and then he fed the goats.
Critical	Ana and Liz joined a pottery club.	Ana picked up some clay with Liz and then she made a vase.
Critical	Liz and Ana planned a fun outing at the lake.	Liz went kayaking with Ana and then she lost her oars.
Critical	Ana and Liz were gardening.	Ana dug a hole with Liz and then she planted a tree.
Critical	Will and Matt went to a corn maze.	Will rode a tractor with Matt and then he drank hot chocolate afterwards.
Critical	Liz and Ana went to a concert.	Liz was dancing with Ana and then she took videos with her phone.
Filler	Ana and Matt were in class.	Ana discussed the lecture topics with Matt and then she took out a pencil.
Filler	Liz and Will went to the pool.	Liz jumped in the water with Will and then she put on her goggles.
Filler	Matt and Liz decided to throw a party.	Matt drove Liz to the store and then he bought lots of balloons.
Filler	Ana and Liz went to Will's new apartment.	Will got a gift from Ana and Liz and then he opened the box.

Filler	Liz and Ana were at marching band practice.	Will gave some water to Liz and Ana and then he picked up the batons.
Filler	Liz and Will joined an orchestra.	Liz practiced with Will and then she accidentally broke her cello.
Filler	Matt and Liz wanted to exercise.	Matt went with Liz to the gym and then he ran on the treadmill.
Filler	Ana and Matt watched the seagulls fly around.	Ana handed Matt a bag and then she scattered crackers on the ground.
Filler	Will and Liz went to the amusement park.	Will handed a ticket to Liz and then he bought cotton candy.
Filler	Ana and Matt were playing tennis.	Ana got the ball from Matt and then she hit it over the net with her racquet.
Filler	Will and Ana were at the train station.	Matt got some money from Will and Ana and then he bought a ticket.
Filler	Ana and Matt went to play soccer.	Ana handed some cones to Matt and then she put on her jersey.
Filler	Will and Liz were bored.	Will wanted to play prank on Liz and then he started hitting her with a pillow.
Filler	Will and Ana went bowling.	Will loaned some shoes to Ana and then he picked up two bowling balls.
Filler	Liz and Matt were doing some chores.	Liz spent all day cleaning the apartment with Matt and then she bought a new vacuum.
Filler	Ana and Matt were pouring drinks.	Ana served some appetizers with Matt and then she took out more cups.
Filler	Liz and Matt were making food in Ana's kitchen.	Liz handed some flour to Ana and Matt and then she put on her apron.
Filler	Ana and Will went fishing.	Ana grabbed a fishing rod from Will and then she got out the bait.
Filler	Matt and Liz were shopping for clothes.	Matt bought a jackect with Liz and then he looked for a hat.
Filler	Will and Ana went to the farmer's market.	Will gave a basket to Ana and then he picked out a tomato.
Filler	Matt and Liz were watching a TV show.	Matt got the remote from Liz and then he turned off the TV.
Filler	Will and Ana were driving home from work.	Will got some coins from Ana and then he paid for the tolls.
Filler	Matt and Ana were getting ready for job interviews.	Matt practiced a speech with Ana and then he put on a tie.
Filler	Matt and Liz were in an art class.	Matt passed some pigments to Liz and the he started painting on a canvas.
Filler	Liz and Matt were building some furniture.	Liz read the instructions with Matt and then she picked some screws.
Filler	Ana and Will were looking at the bookshelves.	Ana got an encyclopedia from Will and then she rearranged the bookends.

Table B2. Stimuli for the main experiment List 1 & 2 (singular exposure)

Item	Story	Story Type
Exposure	Alex went to the store. They looked through all the aisles. Then they bought some milk.	Alex_only
Exposure	Alex wanted to get a treat. The ice cream store was closed. Then they decided to get a cupcake instead.	Alex_only
Exposure	Liz and Alex ordered lunch at the cafeteria. Liz showed the menu to Alex. Then they asked for a sandwhich.	Alex_second
Exposure	Will and Alex were having lunch at the park. Will passed a fancy box to Alex. Then they ate a piece of sushi.	Alex_second
Exposure	Alex was building some furniture. The instructions seemed impossible. Then they realized there was a missing screw.	Alex_only
Exposure	Alex decided to start a band. There were two spots left. They offered to play the xylophone.	Alex_only
Exposure	Alex decided to go for a hike. The trail had a warning about bears. Worried, they bought some bear spray.	Alex_only
Exposure	Alex moved into a new apartment. The air conditioning broke. They decided to buy a fan.	Alex_only
Exposure	Alex was in the mailroom. They picked up the mail. Then they borrowed scissors to open a package.	Alex_only
Exposure	Liz and Alex went to the zoo. Liz wanted to show the antelopes to Alex. But they got distracted by the penguins.	Alex_second
Exposure	Liz and Alex planned a fun outing at the lake. Liz went kayaking with Alex. They lost their oars.	Alex_second
Exposure	Will and Alex went to the amusement park. Will handed Alex a ticket. Then they bought cotton candy.	Alex_second
Exposure	Will and Alex were getting ready to go to the beach. Will packed beach supplies with Alex. Then they realized they forgot their sunscreen.	Alex_second
Exposure	Liz and Alex were getting ready to go shopping. Liz brought the coupons to Alex. Then they wrote a shopping list.	Alex_second
Exposure	Alex was playing video games. Suddenly the power shut off. Then they lit a candle.	Alex_only
Exposure	Alex was scrolling through social media. They realized their phone was on low battery. They plugged it into an outlet.	Alex_only
Exposure	Alex was baking a cake. They mixed all the ingredients in a bowl. Then they poured the batter into a pan.	Alex_only
Exposure	Alex was going to watch a movie. The popcorn was too expensive. Then they bought french fries instead.	Alex_only
Exposure	Alex was watching a TV show. Then the doorbell rang. They paused the show with the remote.	Alex_only
Exposure	Liz and Alex went to a concert. Liz was dancing with Alex. Then they took videos with their phone.	Alex_second

Exposure	Will and Alex were gardening. Will dug a hole with Alex. Then they planted a tree.	Alex_second
Exposure	Will and Alex decided to throw a party. Will drove Alex to the store. They bought lots of balloons.	Alex_second
Exposure	Will and Alex were bowling. Will loaned some shoes to Alex. Then they picked up two bowling balls.	Alex_second
Exposure	Liz and Alex were in an art class. Liz passed some pigments to Alex. Then they started painting on a canvas.	Alex_second
Critical	Will and Alex were at marching band practice. Will gave some water to Alex. Then they put down the batons.	
Critical	Will and Alex went grocery shopping. Will gave the grocery list to Alex. Then they grabbed a bag of chips.	
Critical	Liz and Alex went fishing. Liz handed a fishing pole to Alex. Then they got out the bait.	
Critical	Liz and Alex went to the farmer's market. Liz gave a basket to Alex. Then they picked out some tomatoes.	
Critical	Liz and Alex sorted candy after Halloween. Liz passed a chocolate bar to Alex. Then they ate lollipops.	
Critical	Will and Alex went to play soccer. Will handed some cones to Alex. Then they put on their jerseys.	
Critical	Will and Alex went hiking. Will lent a backpack to Alex. Then they packed some snacks.	
Critical	Liz and Alex were studying together. Liz handed some notes to Alex. Then they looked through the textbook.	
Critical	Will and Alex went to a restaurant. Will handed the salt to Alex. Then they ate pasta.	
Critical	Liz and Alex were at a restaurant. Liz handed the menu to Alex. Then they ordered some wine.	
Critical	Will and Alex were deciding what food to make for a party. Will gave an apron to Alex. They made a cake.	
Critical	Liz and Alex went to the store. Liz gave the cart to Alex. Then they bought some soda.	
Critical	Will and Alex were pouring drinks. Will gave a cup to Alex. Then they had some orange juice.	
Critical	Liz and Alex went to a New Year’s Eve party. Liz gave a drink to Alex. Then they popped party poppers.	
Critical	Will and Alex were at the train station. Will gave some money to Alex. Then they bought two tickets.	
Critical	Liz and Alex joined a pottery club. Liz gave some clay to Alex. Then they made vases.	
Critical	Will and Alex were shopping for clothes. Will showed a cool jacket to Alex. Then they decided to buy hats instead.	
Critical	Liz and Alex went to a bakery. Liz lent some money to Alex. Then they bought croissants.	

Critical	Will and Alex were looking at the bookshelves. Will passed an encyclopedia to Alex. Then they rearranged the bookends.	
Critical	Liz and Alex were cleaning up after a dinner party. Liz handed a towel to Alex. Then they dried the plates.	
Filler	Will and Liz wanted to exercise. Will went with Liz to the gym. He ran on the treadmill.	
Filler	Liz and Will played chess. Liz was losing to Will. He captured the king piece.	
Filler	Liz and Will went to a corn maze. Liz rode a tractor with Will. She drank hot chocolate afterwards.	
Filler	Will and Liz went to a farm. Will saw the cows with Liz. Then she fed the goats.	
Filler	Will was fast asleep. Suddenly his alarm went off. He rushed to find his car keys.	
Filler	Liz went to a bar after work. The day had been exhausting. She ordered a cocktail.	
Filler	Will and Liz were at rehearsal. Will helped Liz set up her music stand. Then he got out the sheet music.	
Filler	Will and Liz were bored. Will played a prank on Liz. She sat on a balloon.	
Filler	Will was getting ready for a job interview. He practiced in front of a mirror. Then he put on a tie.	
Filler	Liz was driving home from work. It was foggy and hard to see. She swerved to avoid a deer.	
Filler	Will and Liz went to the airport. Will went through security with Liz. Then he got some coffee.	
Filler	Liz and Will went to the pool. Liz jumped in the water with Will. Then he put on his goggles.	
Filler	Liz and Will were in class. Liz discussed the lecture topics with Will. Then she sharpened a pencil.	
Filler	Will and Liz went to an art museum. Will admired some photography with Liz. She liked the paintings better.	
Filler	Liz was shopping at a jewelry store. None of the gift sets were affordable. So she only bought a necklace.	
Filler	Liz was taking a flight. The pilot warned about some turbulence. She put on some headphones to distract herself.	
Filler	Liz and Will joined an orchestra. Liz practiced with Will. He broke one of the strings on his cello.	
Filler	Liz and Will were doing some chores. Liz spent all day cleaning the apartment with Will. She loved using the new vacuum.	
Filler	Will was knitting a sweater. It was looking a bit too small. He grabbed some more yarn.	
Filler	Will was folding laundry. A spider crawled nearby. He hit it with a shoe.	

Table B3. Stimuli for the main experiment List 3 & 4 (plural exposure)

Item	Story	Story Type
Exposure	Will and Liz went to the store. They looked through all the aisles. Then they bought some milk.	Will_Liz
Exposure	Liz and Will wanted to get a treat. The ice cream store was closed. Then they decided to get a cupcake instead.	Will_Liz
Exposure	Liz and Alex ordered lunch at the cafeteria. Liz showed the menu to Alex. Then they asked for a sandwhich.	Alex_second
Exposure	Will and Alex were having lunch at the park. Will passed a fancy box to Alex. Then they ate a piece of sushi.	Alex_second
Exposure	Will and Liz were building some furniture. The instructions seemed impossible. Then they realized there was a missing screw.	Will_Liz
Exposure	Liz and Will decided to start a band. There were two spots left. They offered to play the xylophone.	Will_Liz
Exposure	Will and Liz decided to go for a hike. The trail had a warning about bears. Worried, they bought some bear spray.	Will_Liz
Exposure	Liz and Will moved into a new apartment. The air conditioning broke. They decided to buy a fan.	Will_Liz
Exposure	Will and Liz were in the mailroom. They picked up the mail. Then they borrowed scissors to open a package.	Will_Liz
Exposure	Liz and Alex went to the zoo. Liz wanted to show the antelopes to Alex. But they got distracted by the penguins.	Alex_second
Exposure	Liz and Alex planned a fun outing at the lake. Liz went kayaking with Alex. They lost their oars.	Alex_second
Exposure	Will and Alex went to the amusement park. Will handed Alex a ticket. Then they bought cotton candy.	Alex_second
Exposure	Will and Alex were getting ready to go to the beach. Will packed beach supplies with Alex. Then they realized they forgot their sunscreen.	Alex_second
Exposure	Liz and Alex were getting ready to go shopping. Liz brought the coupons to Alex. Then they wrote a shopping list.	Alex_second
Exposure	Liz and Will were playing video games. Suddenly the power shut off. Then they lit a candle.	Will_Liz
Exposure	Will and Liz were scrolling through social media. They realized their phones were on low battery. They plugged them into an outlet.	Will_Liz
Exposure	Liz and Will were baking a cake. They mixed all the ingredients in a bowl. Then they poured the batter into a pan.	Will_Liz
Exposure	Will and Liz were going to watch a movie. The popcorn was too expensive. Then they bought french fries instead.	Will_Liz
Exposure	Liz and Will were watching a TV show. Then the doorbell rang. They paused the show with the remote.	Will_Liz

Exposure	Liz and Alex went to a concert. Liz was dancing with Alex. Then they took videos with their phone.	Alex_second
Exposure	Will and Alex were gardening. Will dug a hole with Alex. Then they planted a tree.	Alex_second
Exposure	Will and Alex decided to throw a party. Will drove Alex to the store. They bought lots of balloons.	Alex_second
Exposure	Will and Alex were bowling. Will loaned some shoes to Alex. Then they picked up two bowling balls.	Alex_second
Exposure	Liz and Alex were in an art class. Liz passed some pigments to Alex. Then they started painting on a canvas.	Alex_second
Critical	Will and Alex were at marching band practice. Will gave some water to Alex. Then they put down the batons.	
Critical	Will and Alex went grocery shopping. Will gave the grocery list to Alex. Then they grabbed a bag of chips.	
Critical	Liz and Alex went fishing. Liz handed a fishing pole to Alex. Then they got out the bait.	
Critical	Liz and Alex went to the farmer's market. Liz gave a basket to Alex. Then they picked out some tomatoes.	
Critical	Liz and Alex sorted candy after Halloween. Liz passed a chocolate bar to Alex. Then they ate lollipops.	
Critical	Will and Alex went to play soccer. Will handed some cones to Alex. Then they put on their jerseys.	
Critical	Will and Alex went hiking. Will lent a backpack to Alex. Then they packed some snacks.	
Critical	Liz and Alex were studying together. Liz handed some notes to Alex. Then they looked through the textbook.	
Critical	Will and Alex went to a restaurant. Will handed the salt to Alex. Then they ate pasta.	
Critical	Liz and Alex were at a restaurant. Liz handed the menu to Alex. Then they ordered some wine.	
Critical	Will and Alex were deciding what food to make for a party. Will gave an apron to Alex. They made a cake.	
Critical	Liz and Alex went to the store. Liz gave the cart to Alex. Then they bought some soda.	
Critical	Will and Alex were pouring drinks. Will gave a cup to Alex. Then they had some orange juice.	
Critical	Liz and Alex went to a New Year's Eve party. Liz gave a drink to Alex. Then they popped party poppers.	
Critical	Will and Alex were at the train station. Will gave some money to Alex. Then they bought two tickets.	
Critical	Liz and Alex joined a pottery club. Liz gave some clay to Alex. Then they made vases.	
Critical	Will and Alex were shopping for clothes. Will showed a cool jacket to Alex. Then they decided to buy hats instead.	
Critical	Liz and Alex went to a bakery. Liz lent some money to Alex. Then they bought croissants.	

Critical	Will and Alex were looking at the bookshelves. Will passed an encyclopedia to Alex. Then they rearranged the bookends.	
Critical	Liz and Alex were cleaning up after a dinner party. Liz handed a towel to Alex. Then they dried the plates.	
Filler	Will and Liz wanted to exercise. Will went with Liz to the gym. He ran on the treadmill.	
Filler	Liz and Will played chess. Liz was losing to Will. He captured the king piece.	
Filler	Liz and Will went to a corn maze. Liz rode a tractor with Will. She drank hot chocolate afterwards.	
Filler	Will and Liz went to a farm. Will saw the cows with Liz. Then she fed the goats.	
Filler	Will was fast asleep. Suddenly his alarm went off. He rushed to find his car keys.	
Filler	Liz went to a bar after work. The day had been exhausting. She ordered a cocktail.	
Filler	Will and Liz were at rehearsal. Will helped Liz set up her music stand. Then he got out the sheet music.	
Filler	Will and Liz were bored. Will played a prank on Liz. She sat on a balloon.	
Filler	Will was getting ready for a job interview. He practiced in front of a mirror. Then he put on a tie.	
Filler	Liz was driving home from work. It was foggy and hard to see. She swerved to avoid a deer.	
Filler	Will and Liz went to the airport. Will went through security with Liz. Then he got some coffee.	
Filler	Liz and Will went to the pool. Liz jumped in the water with Will. Then he put on his goggles.	
Filler	Liz and Will were in class. Liz discussed the lecture topics with Will. Then she sharpened a pencil.	
Filler	Will and Liz went to an art museum. Will admired some photography with Liz. She liked the paintings better.	
Filler	Liz was shopping at a jewelry store. None of the gift sets were affordable. So she only bought a necklace.	
Filler	Liz was taking a flight. The pilot warned about some turbulence. She put on some headphones to distract herself.	
Filler	Liz and Will joined an orchestra. Liz practiced with Will. He broke one of the strings on his cello.	
Filler	Liz and Will were doing some chores. Liz spent all day cleaning the apartment with Will. She loved using the new vacuum.	
Filler	Will was knitting a sweater. It was looking a bit too small. He grabbed some more yarn.	
Filler	Will was folding laundry. A spider crawled nearby. He hit it with a shoe.	

REFERENCES

Arnold, J. E. (1998). Reference Form and Discourse Patterns. Dissertation, Stanford University.

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

Arnold, J. E., Mayo, H. C., & Dong, L. (2021). My pronouns are they/them: Talking about pronouns changes how pronouns are understood. *Psychonomic Bulletin & Review*, *28*(5), 1688-1697. https://doi.org/10.3758/s13423-021-01905-0

Arnold, J. E., Payst, N., Kakoki, E., & Wang, L. (2023). Online comprehension of nonbinary "they" is easier with supporting linguistic context. *Poster presented at the 36th Annual Conference on Human Sentence Processing. Hosted by the University of Pittsburgh, PA*.

Arnold, J. E., Strangmann, I. M., Hwang, H., Zerkle, S., & Nappa, R. (2018). Linguistic experience affects pronoun interpretation. *Journal of Memory and Language*, *102*, 41–54. https://doi.org/10.1016/j.jml.2018.05.002

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*(3), 355–387. https://doi.org/10.1016/0010-0285(86)90004-6

Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, *129*(2), 177–192. https://doi.org/10.1037/0096-3445.129.2.177

Bradley, E. D., Schmid, M., & Lombardo, H. (2019). Personality, prescriptivism, and pronouns: Factors influencing grammaticality judgments of gender-neutral language – corrigendum. *English Today*, *35*(4), 63–63. https://doi.org/10.1017/s0266078419000397

Byrd, C. T., Conture, E. G., & Ohde, R. N. (2007). Phonological priming in young children who stutter: Holistic versus incremental processing. *American Journal of Speech-Language Pathology*, *16*(1), 43–53. https://doi.org/10.1044/1058-0360(2007/006)

Brysbaert, M., Mandera, P., & Keuleers, E. (2017). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, *27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Chambers, K. E., Onishi, K. H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition*, *87*(2). https://doi.org/10.1016/s0010-0277(02)00233-0

Chang, F., Janciauskas, M., & Fitz, H. (2012). Language adaptation and learning: Getting explicit about implicit learning. *Language and Linguistics Compass*, *6*(5), 259–278. https://doi.org/10.1002/lnc3.337

Conrod, K. (2020). Pronouns and gender in language. *The Oxford Handbook of Language and Sexuality*. https://doi.org/10.1093/oxfordhb/9780190212926.013.63

Contemori, C. (2021). Changing comprehenders' pronoun interpretations: Immediate and cumulative priming at the discourse level in L2 and native speakers of English. *Second Language Research*, *37*(4), 573–586. https://doi.org/10.1177/0267658319886644

Contemori, C., Mossman, S., & Ramos, A. K. (2021). Acquiring L2 pronoun interpretation biases: Implicit Learning at the discourse level in L2 English. *Studies in Second Language Acquisition*, 1-8.

Dale, R., Roche, J., Snyder, K., & McCall, R. (2008). Exploring action dynamics as an index of paired-associate learning. *PLoS ONE*, *3*(3). https://doi.org/10.1371/journal.pone.0001728

Farmer, T. A., Cargill, S. A., Hindy, N. C., Dale, R., & Spivey, M. J. (2007). Tracking the continuity of language comprehension: Computer mouse trajectories suggest parallel syntactic processing. *Cognitive Science*, *31*(5), 889–909. https://doi.org/10.1080/03640210701530797

Fine, A., Qian, T., Jaeger, T. F., & Jacobs, R. (2010, July). Syntactic adaptation in language comprehension. *In Proceedings of the 2010 workshop on cognitive modeling and computational linguistics* (pp. 18-26).

Fine, A. B., & Florian Jaeger, T. (2013). Evidence for implicit learning in syntactic comprehension. *Cognitive Science*, *37*(3), 578–591. https://doi.org/10.1111/cogs.12022

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE*, *8*(10). https://doi.org/10.1371/journal.pone.0077661

Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, *42*(1), 226–241. https://doi.org/10.3758/brm.42.1.226

Freeman, J. B., Ambady, N., Rule, N. O., & Johnson, K. L. (2008). Will a category cue attract you? Motor output reveals dynamic com- petition across person construal. *Journal of Experimental Psychology: General*, 137, 673-690.

Freeman, J. B., Pauker, K., Apfelbaum, E. P., & Ambady, N. (2010). Continuous Dynamics in the real-time perception of race. *Journal of Experimental Social Psychology*, *46*(1), 179–185. https://doi.org/10.1016/j.jesp.2009.10.002

Gordon, P. C., & Scearce, K. A. (1995). Pronominalization and discourse coherence, discourse structure and pronoun interpretation. *Memory & Cognition*, *23*(3), 313–323. https://doi.org/10.3758/bf03197233

Jaeger, T. F., & Snider, N. (2008, July). Implicit learning and syntactic persistence: Surprisal and cumulativity. *In Proceedings of the 30th annual conference of the cognitive science society* (Vol. 827812). Austin, TX: Cognitive Science Society.

Johnson, E., Arnold, J. E. (2022). Priming discourse structure guides pronoun comprehension, *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Kaan, E., & Chun, E. (2018). Priming and adaptation in native speakers and second-language learners. *Bilingualism: Language and Cognition*, *21*(2), 228-242. https://doi.org/10.1017/s1366728916001231

Kaiser, E. (2009). Effects of anaphoric dependencies and semantic representations on pronoun interpretation. *Anaphora Processing and Applications*, 121–129. https://doi.org/10.1007/978-3-642-04975-0_10

Kehler, A., & Rohde, H. (2019). Prominence and coherence in a Bayesian theory of pronoun interpretation. *Journal of Pragmatics*, *154*, 63-78.

Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). *Coherence and coreference revisited. Journal of Semantics*, *25*, 1–44.

Kieslich, P. J., Henninger, F., Wulff, D. U., Haslbeck, J. M., & Schulte-Mecklenbeck, M. (2019). Mouse-tracking: A practical guide to implementation and analysis 1. In A handbook of process tracing methods (pp. 111-130). *Routledge*.

Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition*, *107*(1), 54–81. https://doi.org/10.1016/j.cognition.2007.07.013

Langlois, V. J., & Arnold, J. E. (2020). Print exposure explains individual differences in using syntactic but not semantic cues for pronoun comprehension. *Cognition*, *197*, 104155. https://doi.org/10.1016/j.cognition.2019.104155

Leventhal, O., Camilliere, S., Chen, P., & Grodner, D. (2020). Using ERPs to investigate the processing of singular they. *In Poster presented at the CUNY conference on human sentence processing, Amherst, MA (virtual)*.

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

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General*, *118*(1), 43–71. https://doi.org/10.1037/0096-3445.118.1.43

Noll, J., Lowry, M., & Bryant, J. (2018). Changes over time in the comprehension of he and they as Epicene pronouns. *Journal of Psycholinguistic Research*, *47*(5), 1057–1068. https://doi.org/10.1007/s10936-018-9577-4

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, *39*(4), 633– 651. https://doi.org/10.1006/jmla.1998.2592

Prasad, G., & Linzen, T. (2021). Rapid syntactic adaptation in self-paced reading: Detectable, but only with many participants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(7), 1156–1172. https://doi.org/10.1037/xlm0001046

Rohde, H., & Kehler, A. (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition and Neuroscience*, *29*(8), 912-927.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*(4), 606–621. https://doi.org/10.1006/jmla.1996.0032

Segaert, K., Menenti, L., Weber, K., & Hagoort, P. (2011). A paradox of syntactic priming: Why response tendencies show priming for Passives, and response latencies show priming for actives. *PLoS ONE*, *6*(10). https://doi.org/10.1371/journal.pone.0024209

Smith, W. G., Alexis, K., & Kam, C. L. H. (2019). Learning speech-internal cues to pronoun interpretation from co-speech gesture: a training study. *Journal of child language*, *46*(3), 433-458. https://doi.org/10.1017/s0305000918000557