The Frequency of Referential Patterns Guides Pronoun Comprehension

Elyce D. Johnson and Jennifer E. Arnold

Department of Psychology and Neuroscience

The University of North Carolina Chapel Hill

Author note:

We have no known conflict of interest to disclose.

Correspondence concerning this article should be sent to Elyce Johnson.

Email: elycedjohnson@gmail.com

This research was supported by NSF grants 1651000 and 1917840 to J. Arnold. We are grateful to Taylor Steele and Avery Wall for their extensive work on experiments 2a and 2b. We also thank Braxton Hawkins, Darith Klibanow, Heather Mayo, Gustavo Nativio, and Lamar Richards, for their assistance in preparing the stimuli and Qualtrics scripts.

Data and study material can be accessed here: https://osf.io/qb2c3.

Abstract

There is extensive evidence that people are sensitive to the statistical patterns of linguistic elements at the phonological, lexical, and syntactic levels. However, much less is known about how people classify referential events and whether they adapt to the most frequent types of references. Reference is particularly complex because referential tokens can be multiply categorized, raising questions about what can be learned through referential exposure. We test the role of linguistic exposure to referential patterns in five experiments on pronoun comprehension, examining linguistic contexts like "X is doing something with Y" (Exps. 1a,b, and c) and transfer events like "X gave something to Y" (Exps. 2a and b). We ask whether the interpretation of ambiguous he/she pronouns is influenced by recent exposure and find that indeed it is, supporting the hypothesis that adaptation affects discourse processing. In Exp. 1, we further ask whether adaptation persists across three types of referring expressions (he/she pronouns, I/you pronouns, and names) and find that it is limited to he/she pronouns. In Exp. 2, we test whether people can learn both syntactically conditioned and semantically conditioned frequency patterns with transfer verbs. Results showed that they learned both patterns. These results provide critical new evidence that discourse processing biases are informed by exposure to referential patterns.

Keywords: pronoun comprehension, adaptation, priming, sentence processing

The Frequency of Referential Patterns Guides Pronoun Comprehension

A key part of understanding language is identifying the connections between ideas, for example, understanding who *he* refers to in *Matt went to the store with Will. He bought some bread*. However, pronouns like *he* and *she* are underspecified and often ambiguous, raising questions about how people do this. Decades of research have demonstrated that comprehenders use the linguistic and social context to guide interpretation. Yet, there is little consensus about the origin of these linguistic biases. In this study, we test the hypothesis that participants tend to follow the more frequent patterns of reference that they encounter in the input, all other things being equal (Arnold, 1998; 2001; Arnold et al., 2018). If people interpret pronouns differently according to frequency changes in the local context, it would provide direct evidence that they represent these referential frequencies and that they use them to guide pronoun interpretation. We introduce an adaptation task for testing this hypothesis and use this task to explore the types of patterns that can be learned through exposure.

In this study, we focus on two types of linguistic constraints known to affect pronoun interpretation: the subject bias and the goal bias. It is well established that people tend to assume that the pronoun is likely to refer to the grammatical subject (usually the first referent in the sentence) more often than other entities (e.g., Arnold et al., 2000; Gernsbacher, Hargreaves, & Beeman, 1989; Gordon et al., 1993; Jarvikivi et al., 2005; Stevenson et al., 1994). Thus, in the example above, people often assume that *he* refers to *Matt.* This is an example of a syntactic bias, and it contributes to the claim that pronouns tend to refer to topical referents since the subject is a topical position (e.g., Brennan, 1995; Kehler et al., 2008; Rohde & Kehler, 2014; van Rij et al., 2013). In addition, pronoun interpretation is influenced by the semantic structure of the context. For example, verbs about transfer events (e.g., *give* or *get)* describe a transfer of

something from one entity (a source) to another (a goal). In this context, people tend to assign pronouns to the goal. Thus, in *Ana sent a message to Liz. She…* or in *Liz received a message from Ana. She…*, there is an expectation for *she* to refer to Liz, the goal (Langlois & Arnold, 2020; Stevenson et al., 1994).

Our question is whether these linguistic biases can be learned from exposure to the frequency of referential patterns. This idea seems plausible since the contexts that promote pronoun assignment also correlate with a high frequency of reference (Arnold, 1998). For example, pronouns tend to be assigned more frequently and more quickly to subjects than objects or obliques. Similarly, people tend to talk more about subjects than objects or obliques (Arnold, 1998; 2010; Arnold et al., 2018b). Pronouns also tend to be assigned more frequently and more quickly to goals than sources. Similarly, people tend to talk about goals more than sources (Arnold, 2001; Arnold, 2021). This leads to the hypothesis that people can learn about referential frequency from exposure. Suppose a person repeatedly observes that the probability of mentioning a subject or goal is higher than other comparable categories. In that case, they might begin to perceive these positions as signals that subjects and goals are highly expected, thus supporting pronoun comprehension.

A competing hypothesis is that people do not pay attention to the statistics of discourse patterns. There is extensive evidence that the context constrains pronoun comprehension, but contextual constraints may be learned and encoded in ways that are unrelated to frequency. For example, one proposal is that people have representations of discourse categories like "topic" and follow the rule that pronouns should refer to the topic. Indeed, several models suggest that pronouns are preferred for topical referents. For example, van Rij, van Rijn, & Hendriks (2013) propose a computational model that depends on a topic-assignment strategy, as well as

judgments about what the speaker might have said. Centering theory (e.g., Brennan, 1995; Gordon, Grosz, & Gilliom, 1993; Grosz, Weinstein, & Joshi, 2005) suggests that pronouns preferentially refer to the "backward-looking center", which is essentially a formalization of the topic.

Alternatively, people could use contextual constraints to infer the probability that a referent will be mentioned and assign ambiguous pronouns to the more probable referent. This idea is explicit in Bayesian models of reference comprehension (e.g., Frank & Goodman, 2012; Kehler & Rohde, 2013; Hartshorne et al., 2015), which explain the probability of a word (e.g., a pronoun) referring to a particular referent in terms of both the likelihood of the word occurring and its appropriateness for the referent. Much of the support for these models comes from evidence that the semantic context constrains pronoun interpretation. For example, in *Ana admired Liz because she…,* Liz is judged to be the more probable referent because she is the likely cause of Ana's admiration. This is an example of implicit causality, which has been shown to influence the interpretation of ambiguous pronouns (e.g., Garvey & Caramazza, 1974, Johnson & Arnold, 2021; Kehler et al., 2008; Stevenson, Crawley, & Kleinman, 1994).

Critically, work within the topicality and Bayesian tradition does not answer our question about frequency. In principle, either topicality or semantic constraints could be independent of frequency. Topicality is an assessment of what a discourse or sentence is about (Givon, 1983; Reinhart, 1982). Semantic constraints like implicit causality involve judgments about the causal connections between events. Presumably, both of these constraints could be generated on the basis of a semantic representation, without any representation about frequency (see Hartshorne et al. 2015 for an example of this). On the other hand, it is also plausible that referential frequency could inform these categories. People might learn that topics are things that get referred to

frequently (Givon, 1983) and that implicit causes are the referents that get referred to frequently within causal contexts (Guan & Arnold, 2021). In sum, evidence for these constraints does not provide unambiguous support for the exposure hypothesis, but they are not incompatible with the exposure hypothesis either.

One way to explore whether exposure matters is by examining how people's lifetime experience with language relates to their ambiguous pronoun interpretations. Several studies (Arnold et al., 2018; Langlois & Arnold, 2020; Johnson & Arnold, 2021) have done so by using the Author Recognition Task (ART; Stanovich & West, 1989; Acheson et al., 2008; Moore & Gordon, 2015) as a proxy measure for how much print exposure someone has had over their lifetime. In the ART, participants identify author names from a selection of both real and fake author names. Participants' performance on the task is calculated by subtracting the number of incorrect names from the number of correct names they identify. ART scores tend to correlate with relevant indicators of language skill, such as vocabulary skill (e.g., Stanovich & West, 1989; Cunningham & Stanovich, 1991, 1997; Moore & Gordon, 2015), so it is regarded as an approximate measure of linguistic exposure.

Arnold et al. (2018) examined how participants' ART scores relate to pronoun comprehension. Participants watched videos in which a woman told short stories such as *Panda Bear is having pizza with Puppy. He wants the pepperoni slice*. Participants learned the characters' genders at the start of the experiment, and in the critical sentences, the two puppets always had the same gender (here, male). Participants demonstrated their pronoun comprehension by answering questions, e.g., *Who wants the pepperoni slice?* People tended to select the subject character (here, Panda Bear), but critically the rate of subject responses was

higher for people with higher ART scores. Arnold et al. (2019) demonstrated a similar effect with children.

The stories used by Arnold et al. (2018) always followed the structure X is doing something with Y, which we call the "joint action" structure. The two characters fill the same semantic role, which means that the primary difference between the two is syntactic, such that the subject character is clearly the topic of the sentence. Langlois and Arnold (2020) instead examined pronoun comprehension in the context of transfer verbs, where half their stimuli used a goal-source verb (e.g., *Ana received a text from Liz…She…*) and half used a source-goal verb (*Ana sent a text to Liz…She…).* This study tested whether ART scores correlate with both the syntactic subject bias and the semantic goal bias. Yet, while pronoun comprehension exhibited both subject and goal biases, the ART scores only correlated with a preference for the subject, following the same pattern as the joint action contexts.

Thus, these studies together suggest a relationship between exposure and discourse processing, based on the assumption that ART scores are related to amount of reading exposure. This finding suggests that amount of exposure matters, but it does not indicate whether it is reading exposure per se or both spoken and written exposure together. However, one problem with this approach is that it only yields correlational results, raising questions about how exactly exposure influences comprehension biases.

Testing exposure with an adaptation paradigm

In the current study, we turn instead to experimental manipulation of referential frequencies. If people track referential frequencies, local context changes could alter pronoun comprehension biases. That is, people might adapt to the most locally relevant statistics about pronoun reference. We do this by manipulating participants' exposure to a particular **pattern** of

reference. For example, imagine reading a sequence of sentences like in Table 1. In the subject-exposure context, people read numerous stories where the pronoun always refers to the subject, whereas in the nonsubject-exposure context, all the pronouns refer to the nonsubject. Do these patterns affect how people then interpret ambiguous pronouns, like the sentence about Matt and Will above?

There is extensive evidence that people can adapt to linguistic patterns within an experiment, but most of this evidence comes from smaller linguistic units. At the syntactic level, exposure to linguistic structures affects both short-term changes in processing and long-term changes in language learning, e.g., through priming effects (e.g., Arai et al., 2007; Branigan et al., 2005; Chang et al., 2000, 2006; Tooley et al., 2009; 2014; Thothathiri & Snedeker, 2008a,b; Traxler & Tooley, 2008; Traxler, 2008a, b). The frequency of words and structures also guides lexical representation (MacDonald, 2013; Seidenberg & MacDonald, 1999, 2018) and syntactic ambiguity resolution (e.g., MacDonald, 1993; Pickering & Garrod, 2007, 2013; Tanenhaus & Trueswell, 1995). Print exposure also guides individual differences in syntactic processing (Farmer et al., 2016; James, Fraundorf, Lee, & Watson, 2017). Brief exposure can also facilitate processing rare syntactic constructions for adults (Fine & Jaeger, 2013; Fine et al., 2013; Langlois, 2021; Wells et al., 2009), although these patterns can be fragile (Harrington Stack et al., 2018). At the phonological level, statistical learning can guide segmentation patterns (Saffran et al., 1996a, 1996b; Chambers, Onishi, & Fisher, 2003). Participants can also semantically adapt to the designation of uncertainty expressions (Schuster & Degen, 2020), and adaptation may facilitate reading times for rare coherence relations over time (Hoek et al., 2021).

In addition, there is initial support for the idea that people can represent referential structures from pronoun priming studies. Kaiser (2009) tested pronoun interpretation in

sentences with novel verbs like *Stephen tulvered Peter and Diane churbited him*. Response to the question "Diane churbited… (Stephen / Peter)" demonstrated an overall bias toward the object character, following a known syntactic parallelism effect (e.g., Chambers & Smyth, 1998). These sentences were preceded by primes where the pronoun referred to either the subject, the object, or was a simple intransitive (*William swooked Betty and Kevin brucked him/her/Ø*). The object bias was greatest following an object prime, suggesting that people represent abstract anaphoric relations. However, this paper does not report method details or statistics, so the strength of the effect is difficult to assess.

Similarly, Contemori (2019) tested both native and L2 speakers of English on the comprehension of a sentence like *John met Paul while he was in high school,* where comprehension was measured by responses to the question "Who was in high school?" Participants generally preferred the subject but were about 3-5% more likely to choose the object when the sentence was preceded by an object prime, e.g., *Emily liked Brian because he was a good person*. Contemori (2019) also reported an order effect, such that the priming effect got stronger by the end of the task. However, without any control condition, it is impossible to tell whether this effect reflects adaptation over time to the prime structure or simply changes over time in how participants approach the task, perhaps due to fatigue. Together, the Kaiser (2009) and Contemori (2019) studies show that people can represent abstract anaphoric relations, and that people tend to follow previously-encountered structures at the discourse level. Critically, this demonstrates that discourse-level structures have psychological reality and can be encoded in memory.

Other preliminary evidence comes from a study where 5-6-year-old children were trained to follow a first-mention bias for pronoun comprehension (Goodrich Smith, Black, and Hudson

Kam, 2019), although, in their stimuli, this was a linear category and not aligned with grammatical subjecthood in all items. Children heard stories like *Annie and Sarah are having a picnic in the park. They have a lot of food with them. Annie is carrying the picnic basket, and Sarah has a blanket to sit on. She's excited about the cookies*. (Goodrich Smith & Hudson Kam, 2015). At pre-test, children interpreted the pronoun as referring to the first character about 50% of the time. Then they were exposed to stories with disambiguated references for five days of training, after which they were more likely to link the pronoun with the first-mentioned character. Disambiguation was achieved with either a) a gesture, b) an unambiguous pronoun (including two characters of different genders), or c) a name. The effect of training was only significant in the gesture condition. However, the effect was in the same direction in the other conditions, and there was no interaction between training and condition. This study shows that exposure, over time, can change pronoun strategies in young children.

Despite these initial findings, relatively little is known about whether people can track the frequency of referential patterns, and if so, how they do it. While it is well established that people can adapt to phonological, lexical, and syntactic structures, much less is known about what information is represented about higher-level relations at the discourse level. One complication at the discourse level is that every input could contribute to multiple possible conclusions about the frequency of discourse patterns. Consider, for example, this excerpt:

"Isaiah Renfro, a top freshman wide receiver at the University of Washington, was at **his** home in South Los Angeles. **He** had to leave in the morning for spring practice, which was about to start in Seattle. But **he** could tell: Another storm was coming, a gale of anxiety and depression." (Streeter, Nov. 15, 2018, New York Times).

The bolded pronouns unambiguously refer to Isaiah Renfro, providing an opportunity to learn about the types of referential patterns that are most frequent. However, that requires categorizing the input in some way, and multiple categorizations are possible. One question is whether learners treat different types of referring expressions differently. Learners might make **word-based** generalizations, e.g., about the word "he," such that this excerpt provides three exemplars of "he" referring to the subject of the previous sentence. Alternatively, they might generalize around **word categories,** such as pronouns (generalizing across *he, she, I, you*, and other pronouns), or they might group **all references** together, such that the same learning would occur if this passage instead referred to Isaiah Renfro using "Mr. Renfro" and "Isaiah."

A second question is how people categorize the pronoun-antecedent relation. Recall that pronoun comprehension can be guided by both syntactically- and semantically- conditioned categories. The subject bias requires observing that a particular grammatical role tends to get re-mentioned. In contrast, the goal bias requires noticing that a particular thematic role tends to get re-mentioned. For example, consider *Ana got a present from Matt and she opened it in front of everybody*. From this exemplar, people might learn that pronouns tend to refer to subjects, or that pronouns tend to refer to goals, or both. Do people track both of these relationships?

Experiment 1a, b, and c test our first question, namely whether people can adapt at all to referential patterns at the discourse level. It does so by testing whether the frequency of subject references affects pronoun comprehension in joint action contexts, which carry few semantic constraints. It further tests whether people generalize referential expressions based on words, word categories, or all references by examining three types of referring expressions: a) third-person singular pronouns (she, he), b) third-person names (e.g., Ana, Matt), and c) first- and second-person pronouns (I, you).

Experiments 2a and b further test whether people track frequency by examining he/she pronoun comprehension when the frequency of subject references and the frequency of goal references in transfer events are manipulated. This tests whether people can learn the frequency of one discourse dimension while ignoring the other. This further provides a critical test of whether referential frequency is a plausible factor in both syntactically-driven context effects (like the subject bias) and semantically-driven context effects (like the goal bias).

Given the complexity of anaphoric relations, we expect that learning the frequency of discourse patterns should require a large quantity of input, as is obtained over a lifetime. In addition, previous priming techniques have yielded very small effects of pronoun priming (Kaiser, 2009; Contemori, 2019). This suggests that if people shift their processing due to local frequency patterns, they would be most likely to do so in the context of robust evidence. For this reason, we introduce a discourse adaptation task in which people are exposed to numerous stories that consistently exhibit the same discourse pattern, under the assumption that a large set of tokens with a consistent referential structure would increase the chance of observing robust referential adaptation.

Experiment 1

Our first set of experiments examines pronoun comprehension in the context of "joint action" predicates, taking the form "X did something with Y… pronoun…." It is well established that adults preferentially assign pronouns to the subject in this context (e.g., Arnold et al., 2018a; Arnold et al., 2007; Nappa & Arnold, 2014). In addition, there are no semantic differences between the roles of the characters in the first sentence, making it an ideal place to start to assess the relationship between frequency and the subject bias.

We introduce a discourse adaptation paradigm where people are exposed to unambiguous references, e.g., *Matt went to dinner with Liz. She/He ordered some spaghetti.* The survey starts with 20 of these exposure sentences, establishing the local frequency as either 100% subject-reference or 100% nonsubject-reference. Participants then see 12 critical test items with ambiguous pronouns intermixed with another 20 exposure stories. Thus, we expect the local frequency to be established before any critical items are encountered, yielding a contrast between exposure conditions.

This set of experiments first tests whether our adaptation paradigm reveals any sensitivity at all to the local frequency of discourse patterns. Our second question concerns the specificity of these frequencies. Our dependent measure is always the interpretation of ambiguous he/she pronouns. Is this process influenced by the frequency of all reference patterns or just the frequency of pronoun references?

We test this by using different exposure sentence types across the three experiments. Experiment 1a uses unambiguous pronouns, as in the story about Matt and Liz above. Experiment 1b uses the same sentences but replaces the anaphoric pronoun with a name (*Matt went to dinner with Liz. Liz/Matt ordered some spaghetti.)* Theoretically, both patterns might provide evidence about the most frequent reference type. If the subject frequently gets mentioned, even with a name, people might learn that subject mention is common, facilitating the comprehension of both pronoun and name anaphors (Arnold, 1998, 2010). On the other hand, pronouns and names are processed differently, where pronouns may be linked to the discourse context more quickly than names (Ariel, 1990; Brennan, 1995; Gordon & Hendrick, 1998). If so, people may track the frequency of pronoun references differently from the frequency of name references.

We also ask whether third-person and first/second-person references are processed similarly. Experiment 1c tests this by replacing the names and pronouns in the exposure sentences with I/you pronouns (*I went to dinner with you. I/you ordered some spaghetti.)* There are two reasons to suspect that I/you pronouns might be categorized differently. First, these pronouns find their referents in the communicative context, namely the speaker/writer and the hearer/reader. Since interpreting *you and I* does not depend on the discourse context, people may ignore the grammatical function of the last mention of *I* or *you*. Second, even if people keep track of abstract discourse relations, they may additionally represent the lexical items used. The priming of syntactic structures is enhanced by a "lexical boost" when the same words are used in prime and target structures (e.g., Chang, Dell, & Bock, 2006; Traxler, Tooley, & Pickering, 2014). Thus, perhaps priming from I/you exposure sentences would have less of an effect on the interpretation of he/she target sentences.

Methods

Participants

This experiment was administered via Amazon Mechanical Turk. All participants were native English speakers, at least 18 years of age, and were paid $1.25 for a completed study. The Mturk HIT was only available to participants in the US, Australia, and England. We report data for 360 participants (120 for each experiment)[1]; additional participants were replaced as detailed in the participant inclusion section below.

Participants were automatically prevented from continuing the survey if they did not provide consent, if they rated their English knowledge as "poor" or "good" (the only acceptable

[1] For Exp. 1a there are actually 59 participants in the subject exposure condition and 61 participants in the nonsubject exposure condition. When replacing an excluded participant, we accidently ran the nonsubject list instead of the subject list.

answer was "native"), or if they reported learning English after age 6. Fourteen of the content questions were used as kick-out questions, and participants who answered more than 35% of these questions incorrectly (n=5) could not continue.

Materials and Design

The experiment was designed in Qualtrics and included demographic questions, the Author Recognition Task, and the main task. Demographic questions asked about socioeconomic status, age, sex, gender, ethnicity/race, language experience, and English language proficiency.

The main task was to read short stories and answer simple questions, and consisted of two practice items, twelve target items, and 40 exposure items. In each item, participants read a two-sentence story and pressed a button to continue; then, they answered two questions about the story. In the short stories, the first sentence described an event with two characters, one in the subject position and the other as the object of a with-phrase; for example, *Ana played music with Liz.* The second sentence always began with reference to one of the characters using their name or a pronoun. In the target items, the second sentence always began with a pronoun made ambiguous by gender; for example, *she played the piano.*

Written items were presented on one page, followed by two comprehension questions on the following two pages (each question appeared on a separate page) with two forced-choice answer choices per question. All items (target and exposure) were about two of four characters: Ana and Liz (female); Will and Matt (male). Across all stimuli, characters were mentioned an equal number of times as the first-mentioned character, i.e., the grammatical subject, and the second-mentioned character, i.e., the nonsubject.

Our key manipulation occurred on the filler items, which exposed participants to either all subject-reference patterns or all nonsubject-reference patterns in a between-subject

manipulation. The exposure items had the same structure as the target items, except that the two characters were always of a different gender for Exps 1a and 1b. The second sentence always began with an anaphoric reference to one of the characters; the form of this anaphoric expression varied across the three experiments: Experiment 1a used she/he pronouns; Experiment 1b used names, and Experiment 1c used I/you pronouns. Since the characters had different gender, the reference was always unambiguous, even for he/she pronouns. All the exposure items in a particular list followed the same pattern. For example, in Exp. 1, the subject-reference lists included exposure stories such as *Ana played music with Matt. She played the piano* while participants in the nonsubject-referent list saw exposure stories like *Ana played music with Matt. He played the piano;* see Table 2. Two questions followed each exposure story; 20 items had two content questions, and 32 items had one content question and one reference question that asked about the subject (n= 16) or nonsubject (n= 16) character.

In each experiment, there were 40 exposure sentences. Twenty occurred at the beginning of the experiment, which established the pattern before participants responded to ambiguous items. The other 20 exposure items were intermixed with the ambiguous critical items throughout the rest of the experiment to maintain exposure to the referential pattern. After all critical and exposure items, participants received two comprehension questions. There were two types of content questions across items: object questions asked about the object mentioned in the second sentence, e.g., *What did Matt play*; action questions asked about the action in the first sentence, e.g., *What did they do?* Content questions were designed to be easily answered if participants were paying attention to the sentence. For exposure items, the reference questions were unambiguous by gender; together, content and unambiguous reference questions served as one of our inclusion criteria for analysis (see Participant inclusion criteria section).

The reference question asked about who performed the action in the second sentence; this question measured how participants interpreted the pronoun and constituted our dependent measure for the critical items. Reference questions occurred in one of two formats, either asking about the subject or nonsubject character, e.g., *Did Matt wear the blue hat?* (see Table 3). Note that this question probes whether the participant is willing to accept that interpretation of the pronoun instead of probing the preferred interpretation. For critical items, the reference question was manipulated within items across lists, such that for each list, half the questions asked about the subject and half about the nonsubject. Across lists, all items appeared in both question conditions. This manipulation served two purposes. First, it kept participants from falling into a pattern of always saying yes or no. Second, it pulled participants away from ceiling. In the subject-question condition, there is a strong bias to answer "yes," signaling a subject interpretation. There is more variation in the nonsubject question condition, reflecting the ambiguity of the pronoun and a general bias to say "yes." For the filler (exposure) items, the reference question was manipulated between items, such that half the items asked about the first character and half asked about the second character. Due to the exposure manipulation, for the subject exposure lists, the answer was "yes" to all the subject-character questions and "no" to all the nonsubject-character questions, while the reverse pattern held for the nonsubject exposure lists. For the 12 critical items, all stories were followed by the reference question and an action content question. For the 40 exposure items, 20 had an action content question and a reference question, and 20 had an action content question and an object content question. The order in which answers appeared in the multiple-choice selection (top/bottom) was counterbalanced for all questions, e.g., yes/no answer options appeared equally in each position and whether the

incorrect or the correct answer was on top/bottom for check questions appeared equally in each position. The order of question types was also counterbalanced.

In sum, each experiment followed a 2 (subject vs. nonsubject exposure) x 2 (subject question vs. nonsubject question) design. The reference question manipulation occurred within each list across items, and the exposure manipulation was between participants. In addition, we used a control manipulation of forward vs. backward order to control for item order. All lists began with the same 20 exposure stories in the same order to establish the paradigm, but the following 32 items (12 critical plus 20 exposure) were presented either forward or backward. This led to a total of 8 lists.

Following the main task, participants completed the Author Recognition Task. We used a modified version of this task developed by Peter Gordon's lab (p.c.), which was based on previous versions of the task (Acheson, Wells, & MacDonald, 2008; Moore & Gordon, 2015; Stanovich & West, 1989). This task consisted of an array of 126 names, 63 of which are author names, 63 of which are foils. Participants were asked to select only the names that are author names, and they were instructed not to guess. Results from the Author recognition task were used for exclusion purposes but will not be reported here.

Study materials for both experiments 1 and 2 can be accessed here: https://osf.io/qb2c3. The Institutional Review Board of UNC-Chapel Hill approved this study.

Participant Inclusion Criteria

Fourteen questions were used as attention-check questions during the experiment. These questions were straightforward to answer if the participant was paying attention. If the participant answered more than 35% incorrectly (n = 5), it indicated that they were not being attentive to the

survey. They were automatically dismissed from the Qualtrics survey on the sixth incorrect answer.

We used three criteria for inclusion in the analysis for those who finished. 1) We replaced participants if they made more than 25% errors across all the filler questions, including content questions for both exposure and critical items and the unambiguous reference questions for exposure items. 2) Participants were instructed not to guess on the Author Recognition task; if more than 33% of the names they selected were incorrect, it signaled they were guessing, and we replaced the participant. It turned out that a large number of participants were excluded for this reason, perhaps because it was evident that this section of the task included no attention checks; 3) Participants who volunteered that they had a language disorder in the demographic question were replaced. We re-ran these participants to meet a target of 120 participants in each experiment. Table 4 reports exclusions for all experiments.

Procedure

Participants on Amazon Mechanical Turk (Mturk) were directed to a link for the Qualtrics survey. They were given instructions to keep the Mturk page open because, at the end of the Qualtrics survey, a random code was generated, and they needed to write in this code on Mturk to be paid for their participation. At the beginning of the survey, participants read and agreed to the consent form. They were also informed that there would be several check questions throughout the survey, and if they answered too many of those questions incorrectly, they would be dismissed from the survey without pay. Participants then completed demographic questions, the pronoun comprehension task, and the Author Recognition Task. They received a thank-you message and a random code at the end of the experiment.

Analytical Approach

Data from all experiments were analyzed with a mixed-effects logistic regression using SAS proc glimmix with a binary distribution and a logit link. The dependent measure was whether the participant provided a response that was consistent with a subject interpretation of the pronoun: this included both yes responses to a question about the subject character and no responses to a question about the nonsubject character. Predictors included exposure condition (1 = subject reference; 0 non-subject reference), reference question condition (1 = asked about subject; 0 = asked about non-subject), and the interaction between the two. For any significant interactions, we used estimates to test the effect of exposure (our manipulation of interest) for each critical question condition within the model. All predictors were grand-mean centered (data can be accessed here: https://osf.io/qb2c3). Considering possible effects of nesting, participant and item were included as random intercepts. Random slopes were included for the critical question condition by participant, and random slopes were included for prime condition, reference question condition, and the interaction by item.

Results

In Exp. 1a, as shown in Figure 1, participants were more likely to assign the pronoun to the subject character in the subject reference exposure condition (84.04%, SE = .025) than in the nonsubject reference exposure condition (69.13%, SE = .032). Participants also showed a "yes" bias, such that the rate of subject interpretations was higher for the questions about the subject referent (where the subject-interpretation answer was "yes") than they were for the questions about the nonsubject answer (where the subject-interpretation answer was "no"). These patterns emerged in our model as a main effect of exposure and a main effect of reference question. In addition, we found an interaction between exposure condition and reference question condition.

When we probed this interaction with contrasts, we found that the exposure effect was significant for the subject-reference question and marginally significant for the nonsubject-reference question, see Tables 5 and 6.

Unlike Experiment 1a, in Experiments 1b and 1c, there were no effects of exposure condition, but again we saw a "yes" bias such that the subject interpretation was stronger for the questions about the subject than the questions about the nonsubject; see Figure 1. As shown in Table 5, in both 1b and 1c we observed main effects of the reference question condition but no effect of exposure condition nor any interactions.

This same pattern of results was replicated in two additional sets of experiments with smaller sample sizes (see Appendix A). In addition, we performed the same analysis where we included participants who had only been excluded for their performance on the ART. The same pattern of results was found (see Appendix B).

Discussion

Our critical finding is that people can and do track referential frequency, and this frequency can affect how they comprehend ambiguous pronouns. We observed this for Exp. 1a, which used third-person pronominal references (*he* and *she*), as well as in two pilot studies. If the short-term modulations reflect the way we learn a language, it suggests that our discourse processing biases come from our experience with the language. This interpretation extends previous work examining biases in pronoun comprehension (e.g., Arnold et al., 2000; Brennan, 1995; Gordon et al., 1993; Kehler & Rohde, 2008) and suggests that these biases can shift with exposure to an alternate referential pattern. It is also consistent with the hypothesis that the frequency of referential patterns contributes to biases in pronoun comprehension (Arnold, 1998).

The effect of exposure was stronger for the items where the question asked about the subject referent; that is, those items where a subject interpretation of the pronoun results in a "yes" answer, and the effect was only marginal for items where the question asked about the nonsubject referent. The second pilot study showed a similar interaction.

We also observed that people do not draw generalizations from all input. Instead, the interpretation of ambiguous he/she pronouns is only influenced by recent experience with other he/she pronouns. It may not be surprising that exposure to I/you pronouns does not affect the interpretation of he/she pronouns. *I* and *you* refer to discourse participants, which are indexical expressions that refer to the speaker and addressee (e.g., Wechsler, 2010). Thus, a sequence of events (*I played music with you. I played the piano*) can be represented as two links between "I" and the speaker, without any anaphoric relation between the two "I" references. By contrast, interpretation of a third-person pronoun requires identifying the anaphoric link between the pronoun, *she,* and the antecedent, *Ana* (e.g., *Ana played music with Matt. She played the piano).* If I/you pronouns do not generate representations of anaphoric pronoun-antecedent structures, they won't teach people anything useful about the frequency of anaphoric pronouns.

Perhaps more surprising is our finding that pronoun comprehension was unaffected by recent exposure to anaphoric names. In principle, people could learn about re-mention patterns through anaphoric names. For example, they could learn that subjects tend to be re-mentioned by reading sequences like *Ana played music with Matt. Ana played the piano* and *Liz went to a supermarket with Matt. Liz bought some bread.* Yet we found that these patterns did not affect pronoun comprehension. There are two possible interpretations of the contrast between name and pronoun exposure stories.

One possibility is that people do not learn about re-mention frequency from proper names. If so, this would have strong implications for the way referential frequency is encoded. There are at least two ways people could track frequency: a) referent re-mention likelihood or b) pronoun referent likelihood. In Exp. 1a, participants read 40 exposure stories where either the subject or nonsubject character was re-mentioned with a pronoun. They might encode this as "Out of all the **subject (or nonsubject) characters** I encountered, 100% were re-mentioned in the next sentence", which is a representation of referent re-mention, or they might encode this as "Out of all **pronouns** I encountered, 100% referred to the subject (or nonsubject) of the previous sentence." Yet in Exp. 1b, the stories with names exhibited the same frequency of re-mention, in that either the subject or the nonsubject was 100% likely to be re-mentioned. If remention probability underlies the exposure effect, we would also expect learning to have occurred for the name stimuli. Indeed, this finding appears to run counter to one aspect of the Expectancy Hypothesis (Arnold, 1998, 2001), which suggested that all forms of reference contribute to representations about referential frequency patterns. If frequency guides referential probability estimates, this finding also fails to support Kehler and Rohde's Bayesian model, which proposes that pronoun comprehension is influenced by the probability of reference, which is calculated over all types of referential expressions.

On the other hand, a second possibility is that people do learn about re-mention frequency from proper names, but only when they process them anaphorically. Unlike pronouns, names can potentially be interpreted as a direct link to a referent without recognizing the anaphoric connection between the two names. Gordon & Hendrick (1998) proposed a model in which name anaphors are harder to process because they require a two-step process whereby the name causes a referent to be instantiated in the discourse representation, and then this referent is

connected to an already-existing representation for that person. By contrast, a pronoun is directly connected to the existing representation. This implies that the link between pronouns and their antecedents may be more strongly represented than the link between name anaphors and their antecedents. Potentially, some readers may not make it through step two and fail to establish the anaphoric link. If so, it may be easier to learn the frequency of anaphor-antecedent structures with pronouns than names. We assume that the strength of learning is proportional to the amount of input, so a short experiment like ours may only provide evidence for the relatively easier patterns of learning – i.e., with pronouns and not names.

While future work is needed to test these possibilities, our current results establish a critical new finding: recent exposure guides pronoun comprehension. Our results also demonstrate that our adaptation paradigm is useful for testing what types of discourse patterns can be learned in a short experiment. By manipulating exposure over multiple stories, we were able to elicit a stronger effect of exposure than has been found in priming studies (Kaiser, 2009; Contemori, 2019). Our next question concerns the types of referential patterns that can be learned. Do people track references in terms of both syntactic and semantic categories? Can they learn to categorize antecedents on one dimension while ignoring variation on the other dimension? We explore these questions in experiment two.

Experiment 2

The literature has established that referents' syntactic and semantic characteristics guide pronoun comprehension. Transfer events offer a demonstration of both biases. In a story like *Ana got the ball from Liz and she….*, Ana is the preferred referent because she is both subject and goal of the transfer event. However, in *Ana threw the ball to Liz and she…*, participants are relatively more equally split between assigning the pronoun to Ana (the subject) and Liz (the

goal). This suggests that people may learn complex generalizations about how pronouns relate to multiple dimensions of each referent in that each transfer story offers evidence about both syntactic and semantic generalizations. Can readers learn the frequency of both types of patterns?

We test this question with our adaptation paradigm by exposing participants to stories, including transfer events. Since Experiment 1 demonstrated that people only learn from he/she pronouns, Exps. 2a and b only included he/she pronouns in the exposure sentences. In Experiment 2a, all the stories included unambiguous pronouns that referred to either the subject or nonsubject character, where the subject was the goal half the time (e.g., *Ana borrowed the book from Matt … she….*) and the source half the time (*Ana threw the ball to Will… he…*). In experiment 2b, we used the same stories but re-combined them in new lists so that the exposure sentences either always referred to the goal or source character, where the goal was the subject half the time and the nonsubject half the time. Thus, for participants to learn the most frequent pattern, they would have to ignore variation on the other dimension (see Fig. 2).

Methods

Participants

We report data from 115 Amazon Mechanical Turk workers who participated in exp. 2a in exchange for $1.50, and 118 workers who participated in exp. 2b in exchange for $2.50.[2] Additional participants were replaced as detailed in the participant inclusion section. The same criteria for continuing the survey were used as in Experiment 1. In addition, due to an error in the script for Exp. 2a some non-native participants were not excluded from participating, so they were paid but not included in the analysis, and 7 participants were excluded from analysis

[2] Based on pre-testing with lab assistants we estimated that experiment 2a would take about 15 minutes, but the actual median time of completion was 24 minutes. Based on this information we increased the pay for experiment 2b.

for Exp. 2b for having participated in a similar study previously. See table 4 for the exclusion report.

Materials and Design

The Qualtrics survey had the same structure as for experiment 1: consent, demographic questions, Author recognition task, main story task. Questions are accessible at https://osf.io/qb2c3.

The main task was very similar to that in Exp. 1. Participants read two-sentence stories about the same four characters as in Experiment 1 (Ana, Liz, Will, Matt) and answered two questions per story, one content and one reference question. Participants began with two practice questions and then began the main task. There were 44 total items in each list, including 12 critical items and 32 exposure items. The critical stories all involved two characters of the same gender and used an ambiguous pronoun to refer to one of these characters. The exposure stories all involved two characters of different gender, and the pronoun was unambiguous by gender.

Our stimulus stories were adapted from Langlois and Arnold (2020), and they differed from experiment 1 by including three clauses. Sentence 1 established the scenario by introducing the two characters in a conjoined noun phrase, where the first-mentioned was always the same as the subject of the second sentence (e.g., *Will and Matt were taking an exam in class.)* Sentence 2 introduced a context event in the first clause, followed by a second clause with a pronoun. Critically, the verb describing the context event was manipulated within items such that either it was either a goal-source verb (*Will borrowed a pencil from Matt and then he began his exam*.) or a source-goal verb (*Will loaned a pencil to Matt and then he began his exam*.) The source-goal verbs were always used in the prepositional structure (and not the double object structure) so as to maintain a parallel syntactic structure as for the goal-source verbs. The stories were identical for the two verb conditions except for the verb and preposition.

For the critical items, the final pronoun-clause was written to be semantically plausible for both characters, so our dependent measure was the participant's final interpretation of who did that action. The two clauses in the second sentence were conjoined by "and then," which signals a "narrative" coherence relation, that is, a description of what happens next (Kehler et al., 2008). This relation is most consistent with a focus on the endpoint of an event (Stevenson et al., 1994) and supports the bias to interpret the pronoun as coreferential with the goal.

The exposure items occurred in four conditions. Verb type was manipulated between items, such that there were 16 source-goal and 16 goal-source stories. For each story, there were two versions that manipulated which pronoun was used in the last clause, such that it referred to either the first or the second character. Like the critical items, the clauses were conjoined with "and then." However, unlike the critical items, the exposure stories used different ending clauses that were most consistent with the intended referent of the pronoun. This served to provide naturalistic and coherent stories, which helped ensure that people correctly interpreted the unambiguous pronoun. For example, following *Liz and Matt were playing with the frisbee outside. Liz snatched the frisbee from Matt,* the subject/goal-reference condition said, *and then she slipped on some mud*, while the nonsubject/source-reference condition said, *and then he decided not to play anymore*. See Table 7 for examples of each stimulus type.

After each story, there were two questions, one reference question and one content question. The reference question was used as our dependent measure for critical items. All other questions were unambiguous control questions. The reference question probed the event in the last clause and asked participants to identify whether a particular character had done that action. Prior to the practice questions, participants were explicitly instructed to answer the questions based on what was explicitly mentioned in the story. They received feedback if they answered

incorrectly on the two practice questions. As in Experiment 1, half the items asked about the subject character, and half the items asked about the nonsubject character. However, unlike Experiment 1, question type was manipulated between items, such that items only appeared with one question type across lists.[3] Example questions are shown in Table 7.

The stories were organized in the survey such that participants read 12 exposure stories initially to establish the discourse pattern for the task. They then saw 12 critical items (6 in the goal-source and 6 in the source-goal conditions) intermixed with an additional 20 exposure stories. The items were rotated across the two verb conditions and assigned to two lists.

Our primary manipulation was the type of pronoun reference established in all 32 exposure stories. In each experiment, there were two types of exposure, manipulated between subjects, such that there was a total of 4 lists per experiment. Experiment 2a compared subject and nonsubject pronoun reference, while Experiment 2b compared goal and source pronoun reference. Thus, in Experiment 2a, the exposure stories in each list displayed either 100% subject or 100% nonsubject reference patterns, where half of the pronoun references in each list were to the goal and half to the source. In Experiment 2b, the patterns were opposite. The exposure stories in each list displayed either 100% goal or 100% source reference patterns. Half of the pronoun references in each list were to the subject, and half were to the nonsubject. In sum, each experiment followed a 2 (subject or goal vs. nonsubject or source exposure) x 2 (goal-source vs. source-goal verb) x 2 (question type). There was no effect of the forward/backward order manipulation in Experiment 1, so we did not include this control for experiment 2. Figures 2 and

[3] There were a few errors in the stories for exp. 2a, but none of these errors impacted our manipulation or analysis. The errors all occurred on the exposure items, and included a) swapping the subject- and non-subject questions across four items on list 1/subject exposure; b) asking about an unmentioned character in the content question, and c) mentioning or asking about a different character in the context sentence or question in one item, and d) reversing the order of the two questions. These variations are noted with the data and study materials, accessed here: https://osf.io/qb2c3.

3 illustrate the experimental design with List 1 for both experiments. List 2 is identical, except the goal-source conditions are source-goal and vice versa.

Procedure

The procedure was the same as for Exp. 1.

Analytical Approach

The dependent measure for Experiment 2a was based on the answer to the referential question. As in experiment 1, we grouped responses that reflected a subject interpretation of the pronoun (yes to subject questions and no to nonsubject questions) and those that reflected a nonsubject interpretation (no to subject questions and yes to nonsubject questions). For Experiment 2b, we again used the responses to reference questions; however, in experiment 2, we grouped those that reflected either a goal interpretation (yes to goal questions, no to source question) or a source interpretation (no to goal questions, yes to source questions).

As in Experiment 1, data were analyzed in SAS proc glimmix with a logit link. Predictors included exposure condition (1 = subject reference; 0 non-subject reference for exp. 2a; and 1= goal condition ; 0 = source condition for exp. 2b), verb condition (1 = goal-source (gs) vs. 0 = source-goal (sg)), reference question condition (1 = asked about subject; 0 = asked about non-subject), and the two and three-way interactions among the predictors. All predictors were grand-mean centered, and maximal random slopes were included.

Results

Figure 4 illustrates the average responses per condition. In Experiment 2a, we observed the expected goal bias, such that subject responses were higher when the subject was the goal (i.e., for the gs verb type) than when the subject was the source (the sg verb type). We also observed an effect of reference question, such that subject responses were more likely for the

subject question than the nonsubject question (as in experiment 1, this reflects a yes bias). Most critically, we found the exposure manipulation also modulated subject responses: subject responses were more frequent in the subject exposure condition than the nonsubject exposure condition. In Figure 4, this is illustrated by the white bars being higher than the black bars across conditions. These patterns emerged in our statistical analysis as main effects of exposure, verb type, and question.

In Experiment 2b, we observed a very similar pattern, but for the goal responses. We again saw a main effect of verb type, but here this reflects a general subject bias: goal responses were more frequent when the goal was the subject (gs verb type) than when the goal was the nonsubject (sg verb type). We also observed a main effect of reference question, such that there were more goal responses for the goal than the source questions. And critically, there was a main effect of exposure: participants were more likely to assign the pronoun to the goal in the goal exposure than source exposure conditions. Again, this is illustrated by higher white than black bars in Figure 4. There was also a marginal interaction between verb type and exposure, and Figure 4 shows that the exposure effect was numerically stronger for the sg verbs. See Table 8 for model output. As in experiment 1, we also performed the same analysis as described above, where we included participants who had only been excluded for their performance on the ART. The same pattern of results was found; see Appendix B.

Discussion

Experiments 2a and 2b again found that recent experience with he/she pronouns influences the interpretation of ambiguous he/she pronouns. These experiments presented a more complex learning context than Experiment 1 because learning the dominant pattern required

participants to ignore one dimension of the input. In exp. 2a, participants learned a syntactically-conditioned effect (e.g., "pronouns tend to refer to the nonsubject") while ignoring the semantic classification of the antecedent. In exp. 2b, they learned instead a semantically-conditioned effect (e.g., "pronouns tend to refer to the source") while ignoring the syntactic dimension.

These results show that, in principle, both the subject bias and goal bias could be learned by observing the most frequent patterns in the context. Given that both of these patterns are indeed frequent in natural language (e.g., Arnold 2001; Arnold et al., 2018b), it is plausible that people may learn these biases through exposure. This is important for two reasons. First, Exp. 2a and 2b show that people are able to pick up on patterns in the context of the noise that comes from the other dimension, which would be important for learning discourse patterns through natural language exposure. Second, it shows that people can categorize input on multiple dimensions, minimally both syntactic roles and semantic roles. Both of these categories are known to be important for pronoun processing and were likely candidates for testing whether the frequency of referential structures guides pronoun comprehension. Further work is needed to identify the range of categories that are relevant for learning discourse patterns.

General Discussion

Our primary goal in this study was to test whether pronoun comprehension is influenced by changes to the local frequencies of referential patterns. We found that indeed it was, in all three experiments that used he/she pronouns in the exposure sentences (Exp. 1a, Exp. 2a, and Exp. 2b). All experiments used contexts with two human characters, one in the subject position and one in the prepositional object position. Experiment 1used the "joint action" structure, while Experiment 2 used transfer verbs. In both cases, comprehension was shifted toward the dominant pattern.

Critically, in Experiment 1, we saw that the exposure effect only occurred when the exposure stimuli were third-person pronouns, but not when they used other referential forms like proper names or first- and second-person pronouns. We considered two possible explanations for this pattern, both of which have theoretical implications. First, this may suggest that people keep track of referential frequency separately for different referring expressions. Pronouns are specialized for reference to contextually prominent referents, while more explicit forms are more appropriate for less prominent referents (e.g., Ariel, 1990; Gundel et al., 1993). This means that classes of referring expressions (e.g., pronouns, names) may have different frequencies of referring to the previous subject. The core question is whether people represent these frequencies separately or together. It seems plausible that people maintain representations of the different specializations for names and pronouns, but they may also draw generalizations about what types of references tend to occur in a form-independent way (see Arnold, 1998, 2010; Kehler & Rohde 2013; Kehler et al. 2008). Our results do not provide support for representations of referential probability that are independent of referential form. This suggests that if such representations exist, they may only emerge with greater input and longer learning periods. We argued that pronouns provide a direct anaphoric link to the antecedent and therefore are likely to support adaptation to referential frequency more than names, which do not provide a direct anaphoric link (Gordon & Hendrick, 1998).

A second possibility is that referential frequency is learned on a lexically specific basis. Thus, comprehension of the words "he" and "she" may be specifically guided by the behavior of these words in discourse. Consistent with this, in Exp. 1c we found that the frequency of I/you pronoun references did not influence he/she pronoun comprehension. However, recent findings from a companion study lend doubt to this interpretation. When we exposed participants to either

only stories with *he*, or only stories with *she*, we observed an equal priming effect on both *he* and *she* critical items (Ye & Arnold, 2022).

In Experiment 2, we further showed that referential frequency could influence the strength of both syntactic and semantically determined patterns, even when this pattern requires ignoring variation on the other dimension. In Experiment 2a, people observed either that pronouns always refer to the subject or that pronouns always refer to the nonsubject, where the subject was the goal for half the stories and the source for the other half. In Experiment 2b, the exposure stories exhibited the opposite pattern: pronouns always referred to either the goal or the source, where the goal was the subject half the time and the nonsubject half the time. In both cases, we observed a main effect of exposure: pronoun interpretation shifted in the direction of the dominant pattern.

Despite our robust effects of exposure, these effects emerged only as a small shift in comprehension preferences over and above the expected effects of the discourse context. Thus, in all experiments, there was a subject bias. In Experiments 2a and 2b, there was also a general preference to assign the pronoun to the goal character.

As a control manipulation, we also asked our critical question in two ways, either asking about the subject or nonsubject character. This determined whether answering "yes" or "no" was consistent with the subject interpretation. People were more likely to say yes than no in general. This suggests that there is noise in pronoun interpretation, especially in short contexts like this. Participants may not have finished resolving the pronoun at the point of reading the question, so the form of the question influenced their answer. An advantage to this metric is that it allows us to probe the participant's meaning-based interpretation, as opposed to using measures of processing difficulty like reading time. Yet responses tend to be at ceiling (e.g. see Nappa &

Arnold, 2014), so the manipulation of question type is helpful in order to introduce more variation in responses. In this context, we observed effects of exposure, in addition to the well-established subject and goal biases.

Conclusion

Our findings support the idea that exposure to frequent patterns affects language processing at the discourse level. This provides some of the first direct evidence that the frequency of referential patterns affects comprehension biases during pronoun comprehension. It also demonstrates that these biases are malleable, and recent exposure can shift preexisting contextual biases. This work extends prior findings that statistical learning guides acquisition and processing at segmental, lexical, and syntactic levels (e.g., Saffran et al., 1996, Redington et al., 1998). However, learning about discourse patterns is more complex than learning about lower-level patterns because it involves identifying the relevant categories for generalization. Nevertheless, we demonstrated that people do extract generalizations about reference patterns, and these generalizations impact future processing.

Our referential exposure effect is also important for current models of reference processing. First, it supports the hypothesis that discourse constraints on reference comprehension are related to the frequency of discourse patterns (Arnold, 1998, 2001). This provides one possible explanation for how people represent the probability of re-mention in Bayesian models of reference (e.g., Frank & Goodman, 2012; Hartshorne et al., 2015; Kehler et al., 2008; Kehler & Rohde, 2013). However, our results also raise questions about how re-mention probabilities are represented. Prior work (e.g., Arnold, 1998; Kehler & Rohde, 2013) has hypothesized that re-mention probability can be calculated in a form-independent way, but we found that in our short experiment, the pattern of anaphoric name references had no effect on

the comprehension of ambiguous pronouns. Instead, we found that in local contexts, he/she pronoun comprehension is only influenced by exposure to other he/she pronouns. This is consistent with either generalization at the **word level** or the **word class level**, but not at the level of all references. Future work is needed to test whether people can generate form-independent representations of referential frequency under more robust learning conditions.

A key contribution of our study is that it provides a paradigm for studying statistical learning at the referential level. Our approach builds on prior evidence that pronoun interpretation can be primed from the immediately preceding story, but these effects tend to be very small (Kaiser, 2009, Contemori, 2019). Our paradigm instead exposes participants to repeated examples of a particular referential structure. This approach elicits large enough effects to test priming effects on top of other known discourse effects, like verb bias (Exp. 2). It thus provides a promising avenue to test further questions about how people represent referential structures in discourse and how they adapt to context-dependent referential frequencies.

References

Arai, M., Van Gompel, R. P. G., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. Cognitive Psychology, 54, 218-250. https://doi.org/10.1016/j.cogpsych.2006.07.001

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

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

Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. Discourse Processes, 31(2), 137-162. https://doi.org/10.1207/S15326950DP3102_02

Arnold, J. E. (2021). Re-mention frequency is higher for goals than sources in transfer events. Technical Report #6. UNC Language Processing Lab, Department of Psychology & Neuroscience, University of North Carolina – Chapel Hill, Chapel Hill, North Carolina.

Arnold, J. E., Castro-Schilo, L., Zerkle, S., & Rao, L. (2019). Print exposure predicts pronoun comprehension strategies in children. Journal of Child Language, 1–31. https://doi.org/10.1017/S0305000919000102

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

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

Arnold, J. E., Strangmann, I., Hwang, H., & Zerkle, S. (2018b). Reference frequency: what do speakers tend to talk about? Technical Report #2. UNC Language Processing Lab, Department of Psychology & Neuroscience, University of North Carolina – Chapel Hill, Chapel Hill, North Carolina.

Branigan, H. P., Pickering, M. J., & McLean, J. F. (2005). Priming Prepositional-Phrase Attachment During Comprehension. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31, 468-481. https://doi.org/10.1037/0278-7393.31.3.468

Brennan, S. E. (1995). Centering attention in discourse. Language and Cognitive Processes 102,137–67. Stanford, CA.

Chambers, K. E., Onishi, K. H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. Cognition, 87(2), B69–B77. https://doi.org/10.1016/s0010-0277(02)002330

Chambers, C. & Smyth, R. (1998). Structural Parallelism and Attentional Focus: A Test of Centering Theory. Journal of Memory and Language, 39, 593-608.

Chang, F., Dell, G. S., Bock, K., & Griffin, Z. M. (2000). Structural priming as implicit learning: A comparison of models of sentence production. Journal of Psycholinguistic Research, 29, 217-229. https://doi.org/10.1023/A:1005101313330

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. Psychological Review, 113, 234-272. https://doi.org/10.1037/0033-295X.113.2.234

Contemori, C. (2019). Changing comprehenders' pronoun interpretations: Immediate and cumulative priming at the discourse level in L2 and native speakers of English. Second Language Research, 0267658319886644.

Cunningham, A. E., & Stanovich, K. E. (1991). Tracking the unique effects of print exposure in children: Associations with vocabulary, general knowledge, and spelling. Journal of Educational Psychology, 83, 264–274.

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. Developmental Psychology, 33(6), 934–945.

Farmer, T. A., Fine, A. B., Misyak, J. B., & Christiansen, M. H. (2016). Reading span task performance, linguistic experience, and the processing of unexpected syntactic events. The Quarterly Journal of Experimental Psychology. https://doi.org/10.1080/17470218.2015.1131310.

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. PLoS ONE, 8, e77661. doi:10.1371/journal.pone.0077661

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

Frank, M.C. & Goodman, N.D. (2012). Predicting pragmatic reasoning in language games. Science, 336(6084), 998

Garvey, C., & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry, 5,*459-464.

Gernsbacher, M. A., Hargreaves, D. J., & Beeman, M. (1989). Building and accessing clausal representations: The advantage of first mention versus the advantage of clause recency. Journal of Memory and Language, 28(6), 735–755. https://doi.org/10.1016/0749-596X(89)90006-5

Goodrich Smith, W., Black, A. K., & Hudson Kam, C. L. (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.libproxy.lib.unc.edu/10.1017/

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. Cognitive Science, 17, 311-347. https://doi.org/10.1207/s15516709cog1703_1

Gordon, P. C., & Hendrick, R. (1998). The representation and processing of coreference in discourse. Cognitive Science, 22(4), 389–424. https://doi.org/10.1016/S0364-0213(99)80045-7

Guan, S. & Arnold, J. E. (Under review). The predictability of implicit causes: testing frequency and topicality explanations. Revision submitted April 2021.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. Language, 274-307.

Harrington Stack, C. M., James, A. N., & Watson, D. G. (2018). A failure to replicate rapid syntactic adaptation in comprehension. Memory and Cognition, 46(6), 864–877. https://doi.org/10.3758/s13421-018-0808-6

Hartshorne, J. K., O'Donnell, T. J., & Tenenbaum, J. B. (2015). The causes and consequences explicit in verbs. Language, Cognition and Neuroscience, 30(6), 716–734. https://doi.org/10.1080/23273798.2015.1008524

Hoek, J., Rohde, H., Evers-Vermeul, J., & Sanders, T.J.M. (2021). Scolding the child who threw the scissors: Shaping discourse expectations by restricting referents. *Language, Cognition and Neuroscience*. 36(3): 382-399.

James, A. N., Fraundorf, S. H., Lee, E.-K., & Watson, D. G. (2017). Individual differences in syntactic processing: Is there evidence for reader-text interactions? Journal of Memory and Language, 102, 155-181. http://doi.org/10.1016/j.jml.2018.05.006.

Järvikivi, J., Van Gompel, R. G. P., Hyönä, J., & Bertram, R. (2005). Pronominal resolution in Finnish: The impact of subjecthood and word order. Psychological Science, 16, 260–264.

Johnson, E. & Arnold, J. E. (2021). Individual differences in print exposure predict use of implicit causality in pronoun comprehension and referential prediction. *Frontiers*, 12. doi: 10.3389/fpsyg.2021.672109

Johnson, E., & Arnold, J. E. (2021). The Frequency of Referential Patterns Guides Pronoun Comprehension. https://doi.org/10.17605/OSF.IO/QB2C3

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

Langlois, V. (2021). Investigating the role of distribution in syntactic comprehension. Ph.D. Dissertation, UNC-Chapel Hill.

Kaiser, E. (2009). Effects of anaphoric dependencies and semantic representations on pronoun interpretation. Discourse Anaphora and Anaphor Resolution Colloquium, 121-129. Springer, Berlin, Heidelberg.

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

Kehler, A., & Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. Theoretical Linguistics, 39, 1–37.

MacDonald, M. C. (1993). The interaction of lexical and syntactic ambiguity. Journal of Memory and Language, 32, 692-715. doi:10.1006/jmla.1993.1035

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

Moore, M., & Gordon, P. C. (2015). Reading ability and print exposure: Item response theory analysis of the author recognition test. Behavior Research Methods, 47(4),1095–1109.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? Trends in Cognitive Sciences, 11, 105–110. https://doi.org/10.1016/j.tics.2006.12.002

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. Behavioral and Brain Sciences, 36, 329–347. https://doi.org/10.1017/S0140525X12001495

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. Cognitive Science, 22(4), 425–469. https://doi.org//10.1016/S0364-0213(99)80046-9

Reinhart, T. (1982). Pragmatics and linguistics: An analysis of sentence topics. Ms, distributed by the Indiana University Linguistics Club.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996a). Statistical learning by 8-month-old infants. Science, 274(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

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

Schuster, S., & Degen, J. (2020). I know what you're probably going to say: Listener adaptation to variable use of uncertainty expressions. *Cognition*, *203*. https://doi-org.libproxy.lib.unc.edu/10.1016/j.cognition.2020.104285

Seidenberg, M. S., & MacDonald, M. C. (2018). The impact of language experience on language and reading: A statistical learning approach. Topics in Language Disorders, 38, 66–83. https://doi.org/10.1097/TLD.0000000000000144

Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. Cognitive Science, 23, 569-588.

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. Reading Research Quarterly, 24, 402–433.

Stevenson, R. J., Crawley, R. A., & Kleinman, D. (1994). Thematic roles, focus and the representation of events. Language and Cognitive Processes, 9, 519–548.

Tanenhaus, M.K., & Trueswell, J.C. (1995). Sentence comprehension. In. J. Miller & P. Eimas (Eds.), Speech, language, and communication, 11, 217–262. San Diego, CA: Academic Press.

Thothathiri, M., & Snedeker, J. (2008a). Give and take: Syntactic priming during spoken language comprehension. Cognition, 108, 51-68. https://doi.org/10.1016/j.cognition.2007.12.012

Thothathiri, M., & Snedeker, J. (2008b). Syntactic priming during language comprehension in three-and four-year-old children. Journal of Memory and Language, 58, 188-213. https://doi.org/10.1016/j.jml.2007.06.012

Tooley, K. M., Traxler, M. J., & Swaab, T. Y. (2009). Electrophysiological and behavioural evidence of syntactic priming in sentence comprehension. Journal of Experimental Psychology: Learning, Memory, Cognition, 35, 19-45. https://doi.org/10.1037/a0013984

Tooley, K. M., Swaab, T. Y., Boudewyn, M. A., Zirnstein, M., & Traxler, M. J. (2014). Evidence for priming across intervening sentences during on-line sentence comprehension. Language, Cognition And Neuroscience, 29, 289-311. https://doi.org/10.1080/01690965.2013.770892

Traxler, M. J. (2008a). Lexically independent priming in online sentence comprehension. Psychonomic Bulletin and Review, 15, 149-155. https://doi.org/10.3758/PBR.15.1.149

Traxler, M. J. (2008b). Structural priming among prepositional phrases: Evidence from eye-movements. Memory & Cognition, 36, 659-674. https://doi.org/10.3758/MC.36.3.659

Traxler, Matthew J., Kristen M. Tooley, and Martin J. Pickering. 2014. Syntactic Priming during Sentence Comprehension: Evidence for the Lexical Boost. Journal of Experimental Psychology: Learning, Memory, and Cognition 40 (4): 905–18. https://doi.org/10.1037/a0036377.

van Rij, J., van Rijn, H., & Hendriks, P. (2013). How WM load influences linguistic processing in adults: A computational model of pronoun interpretation in discourse. Topics in Cognitive Science, 5(3), 564–580. https://doi.org/10.1111/tops.12029

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. Cognitive Psychology, 58, 250–271.

Weschler, S. (2010). What 'you' and 'I' mean to each other: person indexicals, self-ascription, and theory of mind. Language, 86(2), 332-365. https://www.jstor.org/stable/40666323

Ye, Y. & Arnold, J. E. (2022). Learning the statistics of pronoun reference: for words or linguistic categories? Poster to be presented at the 35th Anuual Conference on Human Sentence Processing, Santa Cruz, CA (virtual).

Appendix A

Additional data: Pilot Studies

Here we report two pilot studies of Experiment 1 that were conducted with a smaller number of participants (n=60) than the current study. The main task in pilot studies 1a, b, and c was the same as the current experiment except that they had fewer exposure items (n = 32). Twelve of these were presented in the initial exposure phase, and the remaining 20 were interspersed with the critical items. Results revealed the same patterns reported in the current study, except that we did not find the interaction between exposure condition and reference question condition in pilot study 1a; see Figure A1 and Table A1.

We then conducted pilot studies 2a, b, and c because we wanted to see whether increasing the number of exposure items (n= 40) would enhance the exposure effect. Critically, we found the same effect as in pilot study 1, and as in the current study, we also saw a significant interaction between exposure condition and reference question condition in pilot study 2a; see Table A2 and Figure A2. When we probed this interaction with contrasts, we found that the exposure effect was significant for the subject-reference question but not for the nonsubject reference question, see Table A3.

The current experiment 1 has the same design as pilot study 2 and serves two purposes. The first was to assess the exposure x reference question interaction with a larger sample. In order to do this, we doubled the number of participants to 120. The second purpose was to correct some minor errors in the first two pilot studies, see Table A4.

Table A1

Pilot study 1a, b,& c experiments (32 exposure items)

	Exp. 1a			**Exp. 1b**			**Exp. 1c**		
Effect	Estimate (SE)	t	p	Estimate (SE)	t	p	Estimate (SE)	t	p
Intercept	1.48 (0.28)	5.22	<.0001	2.15 (0.22)	9.72	<.0001	1.56(0.23)	6.55	<.0001
Question	3.11 (0.38)	8.04	<.0001	2.43 (0.39)	6.08	<.0001	2.81 (0.39)	7.19	<.0001
Exposure	1.87 (0.56)	3.32	.0016	.16 (0.44)	0.36	0.7187	.06(0.47)	.14	0.8867
Question x Exposure	.56 (0.77)	0.73	0.4682	-.89 (0.79)	-1.12	0.2667	0.52(0.78)	0.67	.5044

Table A2

Pilot study 2a, b, & c experiments (40 exposure items)

	Exp. 2a			**Exp. 2b**			**Exp. 2c**		
Effect	Estimate (SE)	t	p	Estimate (SE)	t	p	Estimate (SE)	t	p
Intercept	1.53 (0.29)	5.22	<.0001	2.02 (0.28)	7.00	<.0001	1.10(0.19)	5.53	<.0001
Question	2.62 (0.39	6.57	<.0001	3.14(0.44)	7.00	<.0001	2.63 (0.37)	6.97	<.0001
Exposure	1.61 (0.58)	2.74	<.0082	.54 (0.57)	0.94	0.3527	-0.14(0.39)	-0.37	.7113
Question x Exposure	1.60(0.79)	2.01	0.0484	1.02 (0.89)	1.14	0.2575	-0.11(0.75)	-0.15	.8776

Table A3

Estimates for the exposure effect in pilot study 2a for each reference question condition

Estimate Labels	Estimate	Standard Error	DF	t Value	Pr > |t|
exposure effect for Subject critical Q	2.419	0.804	135.90	3.01	0.0031
exposure effect for Nonsubject critical Q	0.811	0.607	63.59	1.34	0.1863

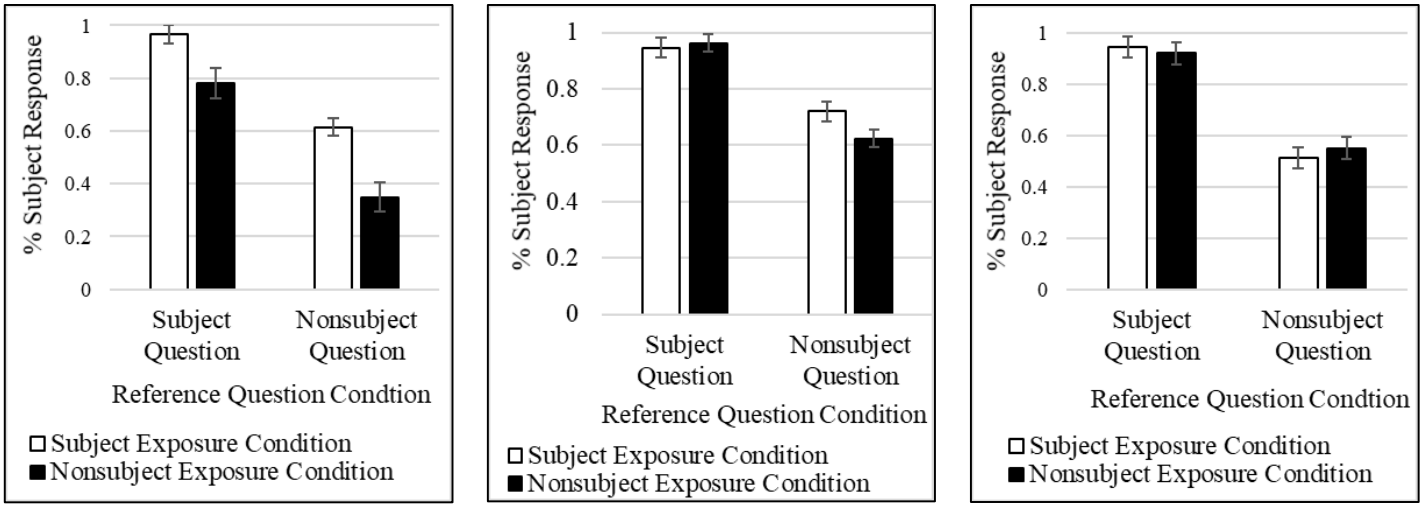
Table A4

Summary of errors in pilot studies 1 and 2.

1. In pilot study 1a and 1b, one item was deleted for the participants who received the backward version for both the subject and nonsubject exposure conditions because the critical question for this item was incorrectly written in the experiment.
2. In pilot study 1c, there was a mistake in one list for participants in the forward version of the subject exposure condition. The question for one item should have asked about the subject, but it asked about the nonsubject.
3. For pilot study 1a-c, there were inconsistencies in the use of determiners for some items (e.g., a cappuccino vs. the cappuccino) and some missing punctuation (i.e., periods and question marks).
4. In pilot study 2a, on one list, for one item, a word was misspelled but it was still understood as intended.
5. In pilot study 2b, on one list, for one item, it said, "Did Ana need…" when it should have said, "Did Ana use…."
6. In pilot study 2c, on one list, for one item, it said, "You was washing the car…" when it should have said, "You were washing the car…."
7. In pilot study 2a-c, on some items, there was missing punctuation (i.e., periods, question marks, and an apostrophe).

Figure A1

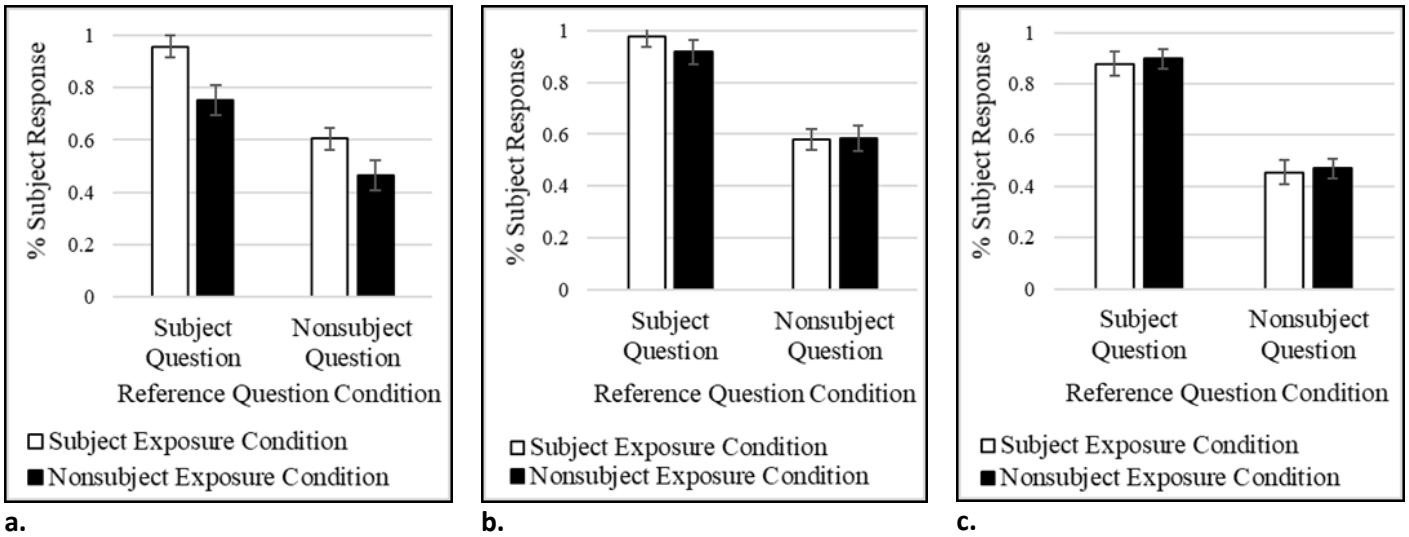
Results of the exposure and question manipulations for Pilot study 1



Note: Error bars depict standard errors of the mean by subject.

Figure A2

Results of the exposure and question manipulations for Pilot study 2



Note: Error bars depict standard errors of the mean by subject.

Appendix B

Including ART Excludes

Here we provide the results for experiments 1 and 2, including all participants who failed the inclusion criterion for the Author Recognition Task and who were not already excluded for some other reason.

Table B1

Results from Exps. 1a, b, and c, including participants who had been excluded for ART only

	Exp. 1a			**Exp. 1b**			**Exp.1c**		
Effect	Estimate (SE)	t	p	Estimate (SE)	t	p	Estimate (SE)	t	p
Intercept	1.70 (0.15)	11.3	<.0001	2.29 (.16)	14.27	<.0001	1.53(.16)	9.21	<.0001
Question	2.22 (0.36)	6.21	<.0001	2.45(0.29)	8.21	<.0001	2.86 (.28)	10.0	<.0001
Exposure	1.57 (0.31)	5.13	<.0001	-.044 (.33)	.31	0.888	.28(.32)	.9	.3690
Question x Exposure	1.14 (0.56)	2.04	0.043	.389 (.55)	0.70	0.4835	.51(.52)	.98	.3298

Note: When those excluded for not meeting the ART criterion are included in the analysis, the total number of participants is 144, 141, and 145 for exps. 1a, 1b, and 1c, respectively.

Table B2

Results from Exps. 2a and b, including participants who had been excluded for ART only

	Exp. 2a			Exp. 2b		
	Estimate (S.E.)	t	p	Estimate (S.E.)	t	p
Intercept	1.42 (0.12)	11.47	<.0001	0.75 (0.14)	5.56	0.0002
Exposure	0.69 (0.23)	3.08	0.0034	0.72 (0.21)	3.42	0.0043
Verbtype	1.1 (0.21)	5.31	<.0001	2.96 (0.3)	9.91	<.0001
Question	1.14 (0.2)	5.56	<.0001	1.57 (0.29)	5.43	<.0001
Exposure X Verbtype	-0.36 (0.36)	-1	0.3253	-0.82 (0.59)	-1.38	0.1776
Exposure X Question	-0.4 (0.35)	-1.15	0.262	-0.5 (0.58)	-0.86	0.398
Verbtype X Question	0.34 (0.38)	0.89	0.3816	0.31 (0.53)	0.58	0.5763
Exposure x Question x Verbtype	0.47 (0.63)	0.74	0.4676	0.36 (0.8)	0.44	0.6663

Note: When those excluded for not meeting the ART criterion are included in the analysis, the total number of participants is 125 and 133 for exps. 2a, and 2b, respectively.

Table 1

Example adaptation paradigm

Subject exposure context	**Nonsubject exposure context**
Liz went to a supermarket with Matt. She bought some bread.	Liz went to a supermarket with Matt. He bought some bread.
Ana went fishing with Will. She borrowed a fishing rod.	Ana went fishing with Will. He borrowed a fishing rod.
Matt rode a bicycle with Ana. He wore a helmet.	Matt rode a bicycle with Ana. She wore a helmet.

Table 2

Example exposure (filler) stories

Exposure Condition	**Subject Reference**	**Nonsubject Reference**
Experiment 1a: He/she pronouns	Ana played music with Matt. She played the piano.	Ana played music with Matt. He played the piano.
Experiment 1b: Names	Ana played music with Matt. Ana played the piano.	Ana played music with Matt. Matt played the piano.
Experiment 1c: I/you pronouns	I played music with you. I played the piano.	I played music with you. You played the piano.

Table 3

Example items and questions

Example Critical Item: Matt went skiing with Will. He wore the blue hat.

Reference subject question: Did Matt wear the blue hat? (Yes / No)

Reference nonsubject question: Did Will wear the blue hat? (Yes / No)

Content (action) question: What did Matt and Will do? (Went skiing / Rode bikes)

Example Exposure Item: Matt wrote a story with Liz. She needed a red pen.

Content (action) question: What did they do? (Wrote a story / Made model airplanes)

Content (object) question: What did Liz need? (A red pen / Sticky notes)

Table 4

Participant exclusions for all five experiments

	Exp. 1a	Exp. 1b	Exp. 1c	Exp. 2a	Exp. 2b
Included in Analysis	120	120	120	115	118
ART guessing	22	21	25	10	15
Accuracy on filler Qs	3		1	5	2
Reported Language Disorder (LD)		2	4		2
Reported Nonnative				1	
Previous participant in a similar study					7
ART and Accuracy	1	3	3	4	3
ART and LD	3	7	9		1
Accuracy and LD		1	1	1	1
LD and nonnative				2	
ART and Accuracy and LD	6	4	6	1	5
Extra participant run by mistake		1			

Table 5

Experiment 1 Results

	Exp. 1a			**Exp. 1b**			**Exp.1c**		
Effect	Estimate (SE)	t	p	Estimate (SE)	t	p	Estimate (SE)	t	p
Intercept	1.86 (0.17)	10.7	<.0001	2.45 (.16)	14.98	<.0001	1.77(.19)	9.18	<.0001
Question	1.95 (0.39)	4.99	<.0001	2.27(0.34)	6.75	<.0001	2.41 (.31)	7.79	<.0001
Exposure	1.5 (0.35)	4.35	<.0001	.1 (.33)	.31	0.7595	.453(.36)	1.26	.2118
Question x Exposure	1.72 (0.6)	2.88	0.0045	.435 (.57)	0.76	0.4455	.033(.55)	.06	.9518

Table 6

Estimates for the exposure effect in Exp. 1a for each reference question condition

Estimate Labels	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
exposure effect for Subject critical Q	2.434	0.526	485.1	4.63	<.0001
exposure effect for Nonsubject critical Q	0.64	0.367	167.7	1.74	0.0834

Table 7

Example stimuli for Experiments 2a and 2b

Target stories
Goal-source condition: Liz and Ana were attending a concert. Liz received the tickets from Ana and then she waited in line. Did Ana wait in line? What did Liz receive from Ana? *Source-goal condition:* Liz and Ana were attending a concert. Liz offered the tickets to Ana and then she waited in line. Did Ana wait in line? What did Liz offer to Ana?
Exposure stories
Goal-source / Subject pronoun: Will and Ana were fishing in the lake. Will retrieved the fishing pole from Ana and then he prepared the bait. Did Will prepare the bait? What did Will retrieve from Ana? *Goal-source / Nonsubject pronoun:* Will and Ana were fishing in the lake. Will retrieved the fishing pole from Ana and then she got a drink from the cooler. Did Will get a drink from the cooler? What did Will retrieve from Ana? *Source-goal / Subject pronoun:* Matt and Liz were cleaning up after a dinner party. Matt gave a stack of dishes to Liz and then he swept the floor. Did Liz sweep the floor? What did Matt give to Liz? *Source-goal / Nonsubject pronoun:* Matt and Liz were cleaning up after a dinner party. Matt gave a stack of dishes to Liz and then she put them in the dishwasher. Did Liz put the dishes in the dishwasher? What did Matt give to Liz?

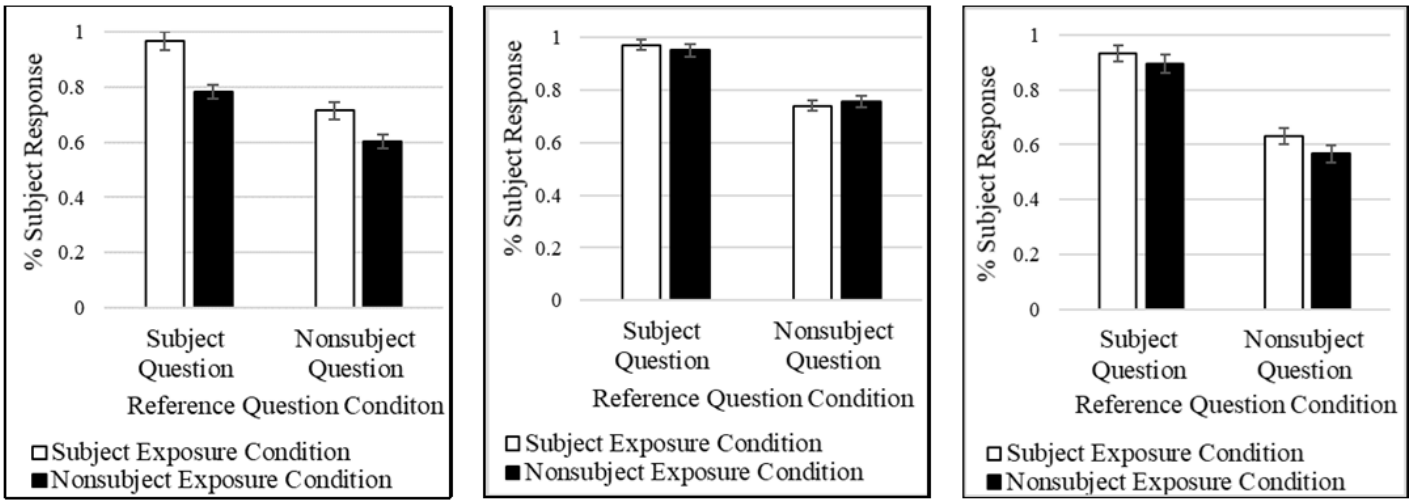
Table 8

Experiment 2 Results

	Exp. 2a (Subj/Nonsubj exposure)			**Exp. 2b (Goal/Source exposure)**		
Effect	**Estimate (SE)**	**t**	**p**	**Estimate (SE)**	**t**	**p**
Intercept	1.52 (0.13)	11.95	<.0001	0.81 (0.16)	5.05	0.0005
Exposure	0.78 (0.25)	3.1	0.0035	0.66 (0.26)	2.56	0.0263
Verbtype	1.19 (0.21)	5.65	<.0001	3.15 (0.31)	10.22	<.0001
Question	1.03 (0.2)	5.02	<.0001	1.35 (0.29)	4.66	0.0001
Exposure X Verbtype	-0.4 (0.42)	-0.96	0.3459	-1.14 (0.59)	-1.92	0.0644
Exposure X Question	-0.39 (0.41)	-0.96	0.3481	-0.66 (0.55)	-1.19	0.2446
Verbtype X Question	0.24 (0.39)	0.61	0.5482	0.29 (0.62)	0.47	0.6463
Exposure x Question x Verbtype	0.36 (0.76)	0.47	0.6434	0.21 (0.98)	0.22	0.832

Figure 1

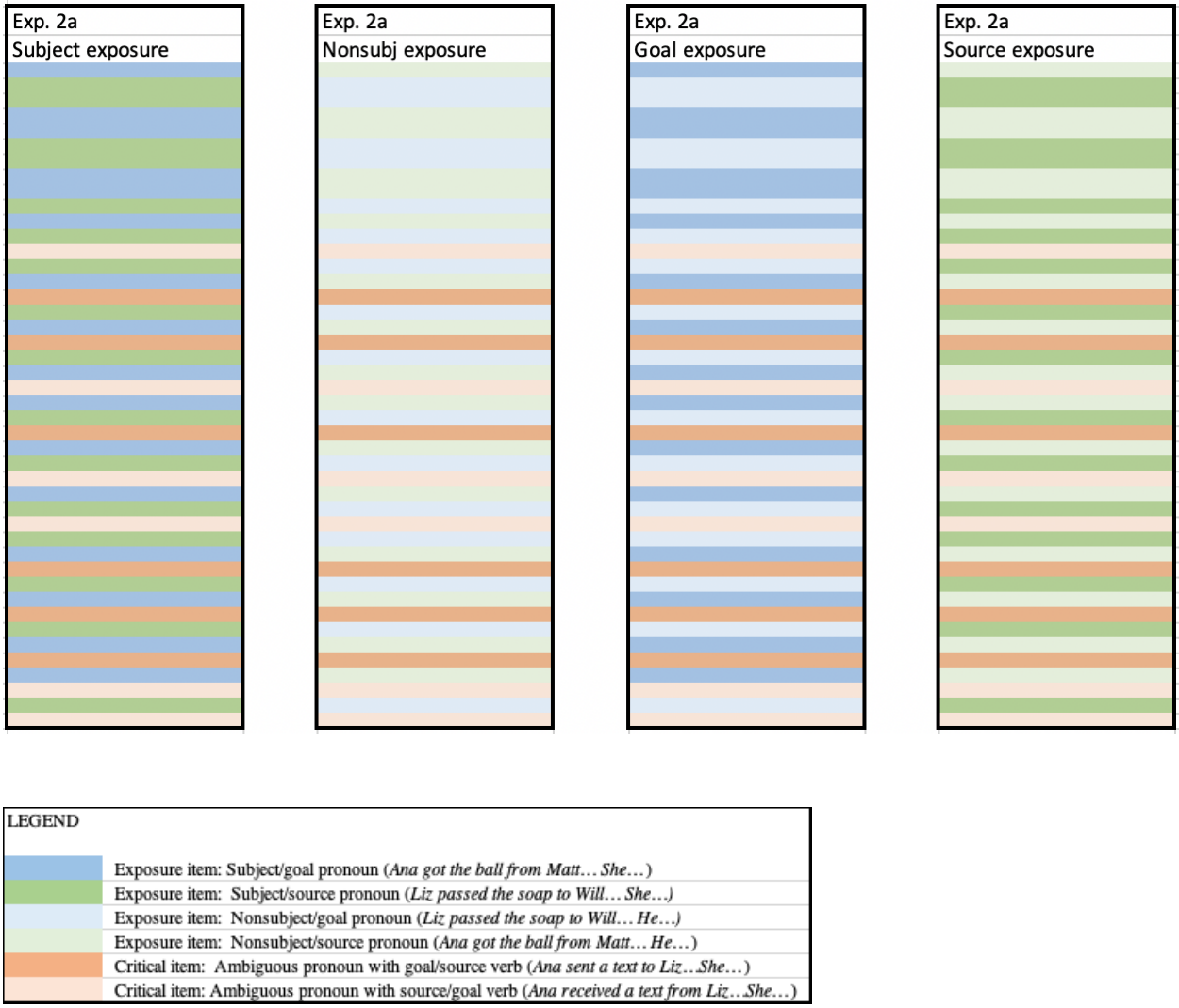
Results of the exposure and question manipulations for Experiment 1



Note: Error bars depict standard errors of the mean by subject.

Figure 2

Illustration of the design of Experiments 2a and 2b.



Note: Each line corresponds with one story and its two questions. Note that the orange lines represent the critical items, and these were the same items across exposure conditions. The exposure conditions (blue and green) were identical except for the reference type.

Figure 3

Illustration of the design of Experiments 2a and 2b.

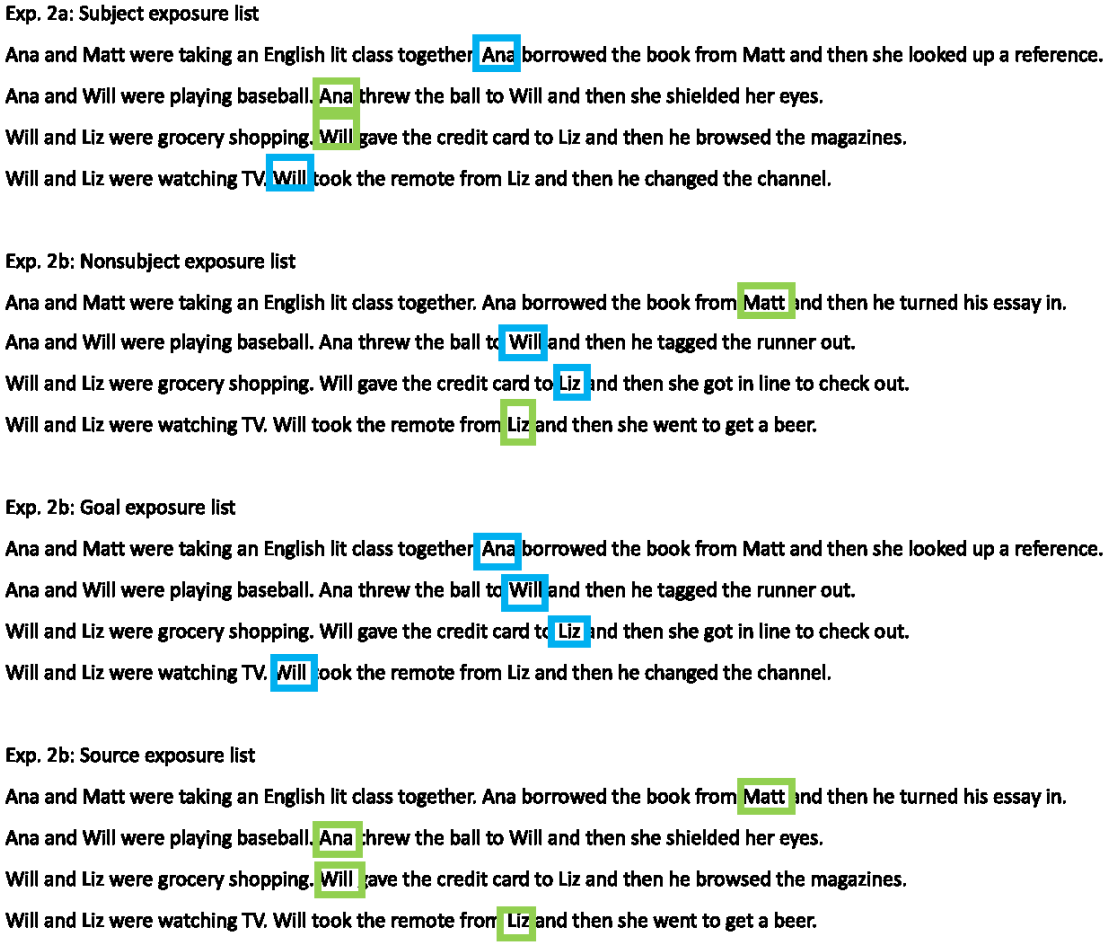
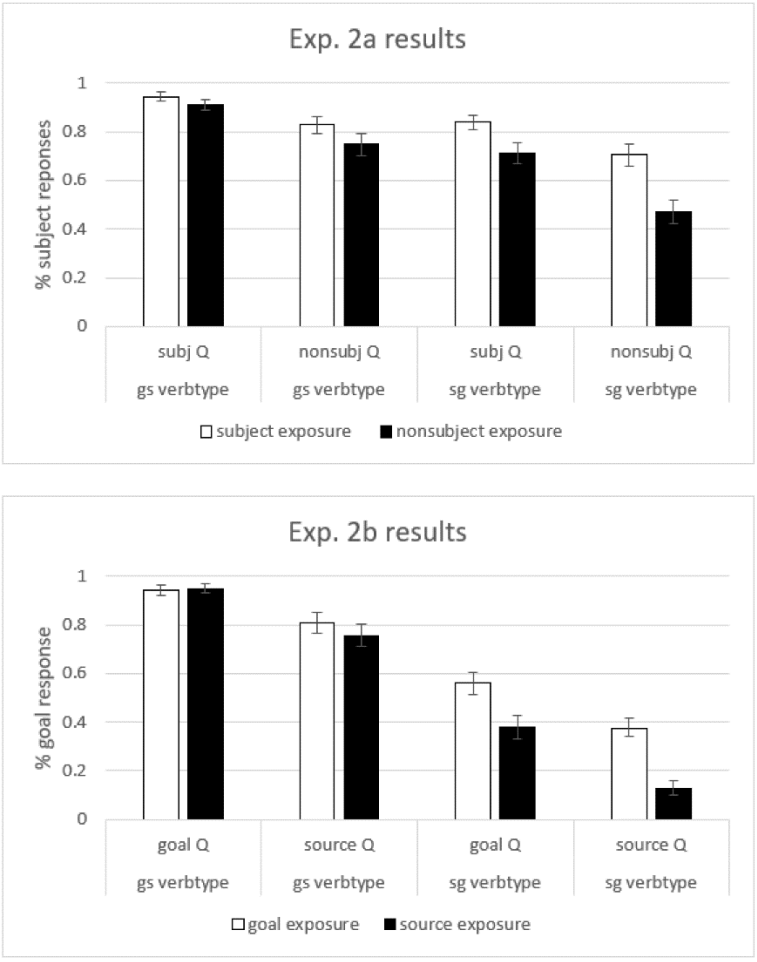


Figure 4

Results of the exposure and question manipulations for Experiment 2.



Note: Top panel: Exp. 2a results, shown as percentage of trials on which the **subject** interpretation of the pronoun was selected. Bottom panel: Exp. 2b results, shown as percentage of trials on which the **goal** interpretation of the pronoun was selected. Error bars depict standard errors of the mean.