Individual differences in print exposure predict use of implicit causality in pronoun comprehension and referential prediction

Elyce Williams[1*], Jennifer Arnold[1]

[1]The University of North Carolina Chapel Hill, Department of Psychology and Neuroscience, Chapel Hill, North Carolina, USA

*** Correspondence:**
Elyce Williams
elyceely@live.unc.edu

Keywords: pronoun comprehension; implicit causality; individual differences; print exposure; sentence processing

ABSTRACT

In three experiments we measured individual patterns of pronoun comprehension (Exps. 1 & 2) and referential prediction (Exp. 3) in implicit causality contexts and compared these with a measure of participants' print exposure (Author Recognition Task; ART). Across all three experiments, we found that ART interacted with verb bias, such that participants with higher scores demonstrated a stronger semantic bias, i.e. they tended to select the pronoun or predict the re-mention of the character that was congruent with an implicit cause interpretation. This suggests that print exposure changes the way language is processed at the discourse level, and in particular, that it is related to implicit cause sensitivity.

Word count: 8739
# of figures: 4
# of tables: 2

1. INTRODUCTION

How does experience with language influence language comprehension? There is extensive evidence that language experience affects lexical and syntactic processing. For example, people tend to produce and understand frequent words and structures more quickly than infrequent words and structures (e.g., Dahan, Magnuson, & Tanenhaus, 2001, Montag & McDonald, 2015; Seidenberg, 1985; Trueswell, 1996; Wells, Christiansen, Race, Acheson, & MacDonald, 2009). However, there is more to be known about how language experience affects processing at the discourse level. In the current study, we explore this further by examining how people interpret ambiguous pronouns, and whether their interpretation is influenced by individual differences in how much they are exposed to written language.

Our project focuses on the role of semantic biases in implicit causality scenarios, namely discourses where causal judgments are important. For example, in *Matt feared Will because he…*, people tend to assume that *he* refers to Will. This interpretation suggests comprehenders are making semantic inferences about who is more likely to be the cause of the fear event. In this example, people tend to expect that the speaker will talk about some action of Will as an explanation of the fearing event, which influences the interpretation of the pronoun *he* (e.g., Kehler & Rohde, 2013). However, this "implicit causality" bias is only one of several constraints known to affect pronoun comprehension. In addition, people tend to interpret the pronoun as co-referential with the grammatical subject (i.e., a subject bias), which here would lead to the assumption that "he" refers to Matt (e.g., Brennan, 1995; Gordon, Grosz, & Gilliom, 1993; Nappa & Arnold, 2014). This means that for any given pronoun, comprehenders must weigh different constraints to judge the speaker's intended meaning.

In this project, we examine whether linguistic experience affects the way comprehenders use implicit causality judgments during pronoun comprehension, and if so, whether it affects early or late pronoun comprehension processes. This question is important because it is well established that individuals vary in linguistic experience and that experience modulates language processing at multiple levels. However, most of this work addresses language processing at the phonological, lexical, and syntactic levels (e.g., Farmer, Fine, Misyak, & Christiansen, 2016; MacDonald, 1993; Saffran, Newport, & Aslin, 1996; Wells et al., 2009). Very little is known about how experience relates to discourse processes like pronoun resolution, and evidence thus far has only demonstrated that print exposure correlates with a syntactically conditioned bias, namely the subject bias (Arnold, Castro-Schilo, Zerkle, & Rao, 2019; Arnold, Strangmann, Hwang, Zerkle, & Nappa, 2018; Langlois & Arnold, 2020). This raises questions about whether language experience can also affect sensitivity to semantic constraints, like implicit causality biases.

In contrast to discourse processing, there is extensive evidence that experience with both written and spoken language affects syntactic processing. There are at least two types of evidence showing this. The first type of evidence is the effect of recent exposure through adaptation and priming. When exposure is manipulated within the context of an experiment, adults can incorporate recent experience with less common structures, exhibiting modified language use according to these patterns. Recent experience facilitates comprehension for less common structures (e.g., Branigan, Pickering, & McLean, 2005; Fine & Jaeger, 2013; Wells et al., 2009; see also Tooley & Traxler 2010), and it biases speakers to choose structures that match those they recently heard (e.g., Bock, 1986; Bock & Lobell, 1990; Branigan, Pickering, Stuart, and McLean, 2000; Ferreira & Bock, 2006; Pickering & Branigan, 1998, 1999; see also

Pickering & Ferreira, 2008). For example, Thothathiri and Snedeker (2008) showed that people tend to interpret temporary ambiguities congruently with syntactic structures they have recently encountered, even for structures that are relatively uncommon in natural language.

A second type of evidence for exposure effects comes from individual difference work. Processing tends to be facilitated for individuals with more language experience or for individuals with exposure to a greater range of words or structures. For example, people with greater exposure to language more easily process and produce low-frequency words and structures than people with less exposure to language (James, Fraundorf, Lee, & Watson, 2018 Montag & McDonald, 2015; Seidenberg, 1985; Wells et al.,, 2009).

Our study instead examines language exposure at the discourse level, focusing on how individuals differ in the mechanisms they use to interpret ambiguous pronouns. Studying individual differences is challenging because experience varies on a number of dimensions. For example, people may differ in either spoken or written language experience, or both. It is difficult to dissociate the effects of the two, which may be correlated. The current study is one of the first to address this question at the discourse level, and because of this, our goal is to focus on a broader question: does at least one type of language experience (here, print exposure) correlate with individual differences in pronoun comprehension?

Our focus is on language comprehension and not reading skills per se. We know that reading practice improves reading skills, but reading also provides experience that influences language processes for spoken language use (e.g., Grolig, 2020; Montag, Jones, & Smith, 2015; Montag & MacDonald, 2015). Thus, our test of pronoun comprehension is through an auditory language task, which tests the hypothesis that print exposure has effects that extend to spoken language comprehension.

In support of this hypothesis, recent findings suggest that people who read frequently exhibit stronger structural biases in pronoun comprehension than people who do not read as frequently. In Arnold, Strangmann, Hwang, Zerkle, and Nappa (2018), print exposure correlated with differences in pronoun comprehension. In this study, a higher score on the Author Recognition Task (ART) was associated with a greater tendency to assign the pronoun to the grammatical subject in a spoken-language comprehension task. In the ART, participants were given a selection of names and asked to identify the real ones, where half of the names were real author names and half were not. Scores on the ART are a gross measure of exposure to literature since a greater recognition of author names is assumed to be correlated with a greater degree of print exposure. This task has been shown to correlate with related measures of reading and reading skills, for example, verbal comprehension, word identification, word naming, reading speed, and vocabulary (Stanovich & West, 1989; Cunningham & Stanovich, 1991, 1997; Moore & Gordon, 2015).

Arnold et al.'s (2018) findings suggest that language exposure is instrumental in developing the subject bias for pronoun comprehension. There are several possible explanations for this. One explanation is that reading increases exposure to the most frequent discourse patterns in language. We know that pronouns frequently refer to subjects (Arnold, 1998; Arnold et al., 2018b). People may learn this frequency through exposure, and reading offers one way to increase quantity of exposure. Reading may also provide exposure to a specific type of language that is helpful for learning this bias, namely language that is internally coherent and decontextualized. A second explanation is that reading may provide practice in drawing inferences from the text, which could facilitate a wide range of inferential mechanisms. We discuss these further in the general discussion.

Arnold et al.'s (2018) finding raises a question about how much people can learn from print exposure. We know that the discourse context constrains pronoun comprehension in multiple ways. Does reading facilitate the use of all aspects of the context or just some? Langlois & Arnold (2020) examined this question by asking whether print exposure affects both syntactic and semantic biases in discourses about transfer events. They examined ambiguous pronoun resolution in sentences using transfer verbs (goal/source verbs). For example, participants saw sentences like: *Ana and Liz were playing basketball. Ana threw the ball to Liz, and then she fell down.* Results showed two simultaneous effects – a bias to assign pronouns to the subject character (Ana) as well as an overall goal bias effect, such that people tended to prefer the goal as the referent of the pronoun. However, the goal effect was not related to individual differences, i.e., there was no relationship between ART score and the goal bias. Instead, results showed the same correlation between print exposure and the subject bias as observed by Arnold et al. (2018). That is, participants with higher ART scores were more likely than those with lower scores to assign the pronoun to the subject, and this trend occurred for both goal-source and source-goal verbs. This seems to suggest that print exposure only affects sensitivity to structural biases. However, this may be due to the relative weakness of the goal bias. Does reading exposure affect the use of stronger semantic biases?

To test this question, here we turn to implicit causality constraints, which are known to strongly influence pronoun comprehension. We test the strength of implicit causality biases by asking people to listen to short passages about an implicit causality scenario, e.g., *Ana and Liz were attending a karaoke party. Ana idolized Liz because she is a great singer*. Our stories use verbs that either put the implicit cause in subject position, as in the previous example (idolize), or in object position (dazzle), e.g., *Ana and Liz were attending a karaoke party. Ana dazzled Liz because she is a great singer.* A question probes their interpretation of the pronoun, e.g., *Who is a great singer?* The rate of selecting the implicit cause character measures the strength of the implicit causality bias for that participant. We ask whether the participant's print exposure (as measured by the ART task) correlates with the implicit causality bias or whether, instead, it correlates with the rate of selecting the subject character.

If print exposure does correlate with either bias, our next question is why it does so. Several models of pronoun comprehension point to the importance of prediction: as people understand language, they generate probabilistic expectations that a referent will be mentioned. If an ambiguous pronoun is encountered, there is a bias toward the expected referent (Arnold, 1998, 2001; Kehler & Rohde, 2013; see also Hartshorne et al., 2015 for a similar idea). These expectations are guided by coherence relations. Coherence relations describe a relationship between two clauses and can influence referential expectations (Kehler et al., 2008, 2013). While the connector word is not necessary to form a coherence relation, the choice of a connector word (e.g., *and then, so, because)* greatly influences the perception of the coherence relation. For example, the expectation for a causal continuation is strengthened when *because* links the first and second clause. For that reason, many implicit causality studies (as well as the current study) make this relation explicit by using the word *because,* which signals an explanation relationship and prompts an expectation to continue speaking about the cause (Au, 1986; Ehrlich, 1980; Kehler & Rohde, 2013; Stevenson et al., 1994). This expectation is integrated with later bottom-up processes driven by the pronoun and subsequent material, which ensure that the pronoun gender and number match the referent and that the post-pronominal material is consistent with the assumed referent (Au, 1986; Brennan, 1995; Brown and Fish, 1983; Corrigan, 2001, 2002, 2003; Gordon, Grosz, & Gilliom, 1993; LaFrance, Brownell, & Hahn, 1997).

A related issue emerges in a debate about the time course of implicit causality effects on pronoun comprehension. The expectation-driven (or focusing) account asserts that verb bias has early effects, such that before any disambiguating information is reached, the implicit cause becomes more accessible and is considered a likely candidate for re-mention (Cozijn, Commandeur, Vonk, & Noordman, 2011; Järvikivi, van Gompel, & Hyönä, 2017; Koornneef & Van Berkum, 2006; Long & De Ley, 2000; McDonald & MacWhinney, 1995; Pyykkönen & Järvikivi, 2010). By contrast, the integration account posits that the implicit cause effect does not occur until later when disambiguating information is integrated (Garnham, Traxler, Oakhill, & Gernsbacher, 1996; Stewart, Pickering, & Sanford, 2000).

While the focus of this debate is the timing of processing, it may be related to the information available at different points in a discourse. Consider a sentence like *Ana and Liz were attending a karaoke party. Ana idolized Liz because she is a great singer.* Readers are likely to assume that Liz is the referent of "she," potentially for two reasons. Liz is the implicit cause of the idolizing event, which may lead readers to anticipate future mention of Liz as soon as they encounter the verb *idolize*, and this prediction would be supported by the word *because*. We consider these "early" sources of information. In addition, the second clause provides a plausible explanation for the idolizing event, namely that someone is a great singer. Given that possessing skills is frequently seen as an explanation for being idolized and not a reason to idolize someone else, this leads to the inference that *she* probably refers to Liz. We consider these inferences to be "late" sources of information.

Our question here is whether print exposure correlates with early expectation-driven processes or only later inference-driven processes. We test this in two ways. First, we use two different tasks for testing print exposure. In the first task (Experiment 1), participants hear short stories where the subordinate clause ends with the novel word "dax," for example, *Ana and Liz were attending a karaoke party. Ana idolized Liz because she is a dax* (task-based on Hartshorne & Snedeker, 2013). They are then asked, *Who is a dax?* and in this example, have a choice between Ana or Liz. In the second task (Experiment 2), participants hear short stories like *Ana and Liz were attending a karaoke party. Ana idolized Liz because she is a great singer.* In this task, they are asked *Who is a great singer?* Our task only collects offline judgments (i.e., after comprehension has finished), which means we cannot pinpoint when the judgments are being made. Nevertheless, by using a novel word (dax), we limit the possible constraints from post-pronominal material. This means we can assume that responses are based mostly on early information, i.e., that which occurs prior to the pronoun. Critically, since neither the pronoun nor the novel word carries any disambiguating information, an implicit cause interpretation would have to be made just on the information available from the verb and the clausal connector i.e. *because*.

Second, in Experiment 3, we test whether print exposure correlates with predictions about who will be mentioned next in the absence of a pronoun. We provide participants only with the initial fragment, e.g., *Ana and Liz were attending a karaoke party. Ana dazzled Liz because …*, and ask them to judge which character is most likely to be mentioned next. If the next mention-judgments follow the same pattern as the pronoun comprehension judgments, it will suggest that the effects stem from information available during early processing prior to the pronoun. Supplementary materials for all experiments are available at http://arnoldlab.web.unc.edu/publications/supporting-materials/williams-arnold/ and data are available at https://osf.io/fuzp2/.

2. EXPERIMENTS 1 & 2

2.1. Methods

2.1.1. Participants

Both Experiments 1 and 2 were administered via Amazon Mechanical Turk. All participants were native English speakers, at least 18 years of age. All participants were paid for participation. Our target sample was 60 participants in each experiment, following the sample size used in other studies that tested the effect of print exposure (e.g., Arnold et al., 2018). Participants were excluded if more than 1/3 of their responses on the Author Recognition Task were incorrect because it signals that the participant was ignoring the instructions to not guess. Second, participants were excluded if they answered fewer than 75% of the filler questions correctly because this signals that they were not consistently paying attention. These measures are especially important given the use of online participant recruitment to restrict our sample to participants who were paying attention. We replaced the excluded participants to meet our target sample size of 60 in each experiment.

In Experiment 1, 21 were replaced (16 for guessing on ART, 3 for low filler accuracy, and 2 for both). In Experiment 2, 29 were replaced (7 for guessing on ART, 7 for low filler accuracy, and 15 for both). Our final sample for analysis was 60 participants in each experiment.

2.1.2. Procedure

On Amazon Mechanical Turk, participants were directed to a link for the Qualtrics survey. At the beginning of the survey, participants were instructed to read and indicate that they agreed to the consent form. They were also informed that there would be several check questions in the survey and that if they had too many incorrect responses, they would be dismissed from the survey without pay. Participants then completed the demographic questions, the main task, and then the author recognition task in that order. At the end of the experiment, they received an end-of-survey message thanking them for their time and a randomly generated number to record on the Amazon Mechanical Turk site for payment.

2.1.3. Materials and Measures

Both experiments were designed in Qualtrics. In the main task, participants heard an audio recording of the story and saw pictures of the mentioned characters. After listening to the audio clip, they answered two multiple choice comprehension questions about the story's content. Following the main task, they completed the Author Recognition Task (ART).

The main task consisted of two practice items and twelve target items. Experiment 1 also had eight filler items, while Experiment 2 had 12 fillers. Sample target stimuli are shown in **Table 1.** For both experiments, stimuli included a context sentence that introduced two same-gender characters in a conjoined subject NP, followed by a target sentence that used a verb with a strong implicit causality (IC) bias, where the two characters fell in subject and object positions, e.g. *Ana and Liz were attending a karaoke party. Ana dazzled Liz because she ….* All target items used same-gendered characters, making the pronoun ambiguous. The two experiments were almost identical, but Experiment 1 used the novel word *dax* in the final clause of all sentences (e.g., *…because she is a dax*), while Experiment 2 used a real ending (*…because she is a great singer).* Thus, in Experiment 1, the final clause was semantically ambiguous with respect to the pronoun identity, whereas in Experiment 2, the final clause was semantically more consistent with the implicit cause interpretation of the pronoun. Stories were followed by two questions; the critical question probed the interpretation of the pronoun (*Who is a dax?* Or *Who is a great singer?).*

The critical items followed the same manipulations and experiment design for the two experiments. The 12 target items appeared in two versions: one in which the second sentence contained a subject-biased implicit causality (IC) verb, and one in which the second sentence contained an object-biased IC verb. Thus, for the 12 target stories, we used a total of 24 verbs, 12 subject-biased, and 12 object-biased. The two versions of the target items were the same in that they shared the same context sentence. Verb type was manipulated within-subject such that each participant heard six object-biased items and six subject-biased items. Thus, two lists were created with one version of an item per list (**Table 1**). As a control variable, there were two versions of each list in forwards and backward order, for a total of 4 lists, for all items see **Appendix A** in Supplemental Materials.

Lists were matched for average verb bias, verb frequency, and verb valence rating. Verbs were selected from experiment 2 of Hartshorne and Snedeker (2013), where verb bias was recorded as the percentage of participants who selected the object as the referent of the pronoun. The verbs selected for this study had either an object bias of greater than 70% (object-biased condition) or an object bias of less than 30% (subject-biased condition). Verbs were also classified according to their placement in the Levin (1993) verb classification in which all verbs are either of the class 31.1 (amuse verbs), which tend to be subject-biased, or of the class 31.2 (admire verbs), which tend to be object-biased. Measures of frequency for the verbs were collected via the SUBTLEX-US word frequency database (Brysbaert & New, 2009), and valence ratings came from Warriner, Kuperman, and Brysbaert (2013). See **Appendix A** in Supplemental Materials for biases, ratings, and frequencies.

Filler items were similar to the target items in that they consisted of two sentences, a context sentence, and a two-clause sentence, but unlike the targets, they were disambiguated by different gender characters, and they were not implicit cause sentences. In Experiment 1, the eight fillers took the form *X is doing something with Y,* which we term the "joint action" structure, e.g., *Liz and Will were on vacation. Liz watched TV with Will because she is a dax.* In Experiment 2, where the final clause was consistent with a single interpretation of the pronoun, we sought to increase variation across responses so that participants would not fall into a pattern of responding. We, therefore, used six joint action fillers, plus another six fillers with transfer verbs, e.g., *Will and Ana were at McDonald's. Will took the fries from Ana because she was full*.

There were two types of questions for both target and filler items. Critical questions always asked, *who is a dax?* (Experiment 1) or, e.g., *who is a great singer?* (Experiment 2). These questions tested how the participant interpreted the pronoun. Critical question answer choices were always the two characters in the sentence, and content question answer choices were always yes/no. The other question was a content question that also functioned as an attention check question for filler items. These questions asked about information in the second clause of the sentence. For example, in the sentences, *Liz and Matt were studying for an exam. Liz went to the library with Matt because he is a dax*, the content question asked *Did Liz go to the library with Matt?* The answer choices were either *yes* or *no*. Both the critical question and the content question were formatted similarly in that they were multiple choice questions with two answer options. The content question responses for 8 of the filler items were used to make sure that participants were reading the sentences; if participants missed more than 25% (i.e., 3 items), they were automatically dismissed from the survey and not paid. In addition, we checked all filler content question responses and replaced any participants who missed more than 25% of the total. For all questions, the order in which answers appeared in the multiple-choice selection (top/bottom) was counterbalanced for all questions so that yes/no answer options appeared

equally in each position and character name options (e.g., Liz/Ana) appeared equally in each position. Whether they received a critical question or a content question first or second was also counterbalanced.

For all items (target and filler), there were four characters: Ana, Liz, Will, and Matt (**Figure 1**). For the images that accompanied the audio, the first-mentioned character always appeared on the left side, and the second-mentioned character always appeared on the right side. Across all stimuli, each of the four characters occurred equally in subject and object position. See **Appendix A** in Supplemental Materials for a transcription of the stimuli. Audio and images were presented on the same page, followed by two comprehension questions on the next page with two forced-choice answer choices per question. Audio recordings auto-played, and the button to advance to the next page did not appear until 5 seconds after the duration of the audio to ensure that participants listened. Each question appeared on a separate page.

Prior to the main task, participants answered background questions about themselves, their language experience, and socioeconomic status. Following the main task, participants completed the Author Recognition Task. We used a modified version of this task developed by Peter Gordon, which was based on previous versions of the task (Stanovich & West, 1989; Acheson, Wells, & MacDonald, 2008; Moore & Gordon, 2015; See **Appendix B** in Supplemental Materials). Participants saw an array of 126 names, 63 author names, and 63 foils. Participants were asked to select only the names that are author names, and they were instructed not to guess. Their ART scores were calculated based on the names they selected. They received 1 point for each correct name (+1) and were penalized for selecting an incorrect name (-1). Too many incorrect name selections (more than 1/3 of the names selected) was used as a metric to identify participants who guessed or selected names without reading; these participants were not included in the analysis.

2.2. Analytical Approach

We analyzed the data 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 selected the grammatical subject as the referent of the pronoun or not. It was coded as a binary measure with 1 for a grammatical subject selection and 0 for no grammatical subject selection. Predictors included verb bias, which was effect coded (.5 = subject-biased; -.5 = object-biased), and ART scores, which were grand-mean centered. All models used random intercepts for both participant and item and maximal random slopes except where noted.

For each experiment, we first performed a baseline model that included only the manipulated verb bias to assess its effect. Then for our primary analysis, we added print exposure scores and their interaction with verb bias.

2.3. Results

2.3.1. Experiment 1 Results

Participants were more likely to identify the subject as the referent of the pronoun in the subject-biased condition (64.4%, SE = .042) than in the object-biased condition (11.4%, SE = .022). The baseline model confirmed that this effect of verb bias was significant ($\beta = 2.987$, SE = 0.350, $t = 8.53$, $p < .001$).

Our primary question of interest was whether individuals' pronoun interpretation would vary by their performance on the author recognition task and whether this variation would be correlated with variation in an implicit cause bias, a subject bias, or both. Participants who scored higher on the ART were more likely to select the subject for the subject-biased verbs, and more likely to select the object for the object-biased verbs, resulting in a greater effect of verb

bias than for participants who scored lower on the ART. (**Figure 2**). In short, participants with greater print exposure showed a stronger implicit causality bias than participants with lower print exposure.

To test the significance of this pattern, grand-mean centered ART scores were added to the model, as well as the interaction between ART and verb bias. Results again showed a main effect of Condition (β = 2.903, SE = 0.333, t = 8.71, p <.001). There was no main effect of ART (p = .702), but there was a significant interaction between ART and Verb bias (β = .091, SE = 0.0261, t = 3.49, p <.001), such that as ART score increased, participants were more likely to follow a semantic bias.

2.3.2. Experiment 2 Results

Again, participants were more likely to identify the subject as the referent of the pronoun in the subject-bias condition (78.3%, SE .032) than in the object-bias condition (8.3%, SE .018). The baseline model again revealed a main effect of verb bias (β = 4.074, SE = 0.397, t = 10.26, p <.001).

As in Experiment 1, we also found that participants with higher ART scores exhibited a stronger implicit causality bias for pronoun comprehension compared with participants with lower ART scores. Grand-mean centered ART scores were added to the model, as well as the interaction between ART and verb bias. Results again revealed a main effect of Condition (β = 4.113, SE = 0.388, t = 10.6, p <.001). There was no main effect of ART (p = .823), but there was a significant interaction between ART and Verb bias (β = .105, SE = 0.029, t = 3.59, p <.001)[1], such that as ART score increased, participants were more likely to follow a semantic bias (**Figure 3**).

Our analyses were aimed at assessing individual differences in the implicit causality bias. Nevertheless, we note that in both experiments, participants did not show a subject bias, as some other studies have reported (e.g., Kehler et al., 2008; Stevenson, Crawley, & Kleinman, 1994). Instead, they showed a general object bias. Across all items, the average subject response was 37.9 % for Experiment 1 and 43% for Experiment 2.

2.4. Experiment 1 and 2 Discussion

Experiments 1 and 2 tested individual differences in the interpretation of pronouns, examining causal contexts where the verb bias is expected to guide interpretations. We found that individual performance on a print exposure task was correlated with individual differences in the pronoun task: participants with higher ART scores were more sensitive to the verb bias than participants with lower ART scores.

Notably, the interaction between ART and verb bias was consistent across Experiments 1 and 2, even though the pronoun-clause was ambiguous in Experiment 1, but not in Experiment 2. In Experiment 1, the nonword "dax" provided no information about the pronoun referent, requiring participants to rely only on the pre-pronoun context (verb bias and "because" connector) for interpretation. In Experiment 2, the post-pronoun context provided a continuation that was semantically consistent with the verb bias, so in theory, participants could use either the pre-pronoun context or the post-pronoun context or both. Our findings suggest that print exposure is correlated with participants' sensitivity to the pre-pronoun context, which was available in both experiments. The fact that we see similar implicit cause effects occur in each experiment suggests that there is something over and above the information in the last clause of the implicit cause sentences that leads to an implicit clause bias. However, we cannot disentangle the importance of verb bias from coherence relation. Our stimuli used the word "because," which prompts an

[1] The ART slope by items was removed from this model because it would not converge.

expectation of information about implicit cause (Au, 1986; Stevenson et al., 1994; Kehler & Rohde, 2013; Guan & Arnold, under review). Thus, the observed individual differences may relate to the use of the verb and/or the use of the coherence relation as signaled by the word "because."

We found no subject bias, that is, no general preference to select the subject character as the referent of the pronoun. The lack of a general subject bias contrasts with other studies with similar stimuli (e.g., Kehler et al., 2008; Stevenson, Crawley, & Kleinman, 1994). Instead, we find an object bias, similar to Hartshorne & Snedeker (2013). This raises an interesting question as to why the historically found subject bias is not present in this study. In any case, in principle we could have still seen individual variation in the degree of **relative** subject bias, yet we did not. Critically, ART scores did not correlate with the subject bias overall, which contrasts with previous findings (Arnold et al., 2018; Langlois & Arnold, 2020). We will take this up in the general discussion.

Together the results of Experiments 1 and 2 showed that people differ from each other in how strongly they follow the implicit causality bias. Similar results across experiments suggest that individual differences stem from information up until the pronoun, but not including the post-pronominal information. This suggests that individuals may differ in the degree to which they use the context to make predictions about the upcoming discourse. Participants who read frequently may be better equipped to use the semantic context to generate predictions about likely upcoming mentions. If frequent readers are more consistent in their expectations that the implicit cause will be mentioned, they would also be likely to transfer that prediction onto the interpretation of the pronoun.

We test this idea explicitly in Experiment 3 with the use of a metalinguistic prediction task. We know that verb bias affects predictions: when people read a fragment like *Ana admired Liz,* they judge that *Liz* should be more likely to be mentioned than *Ana* (Guan & Arnold, under review; Weatherford & Arnold, in press). Our question here is whether this bias varies across individuals. Do frequent readers also show a stronger sensitivity to implicit causality when making predictions about who will be mentioned next?

3. EXPERIMENT 3

3.1. Methods

3.1.1. Participants

Seventy-nine native speakers of English participated through Amazon Mechanical Turk. A total of 19 participants were excluded (2 for incorrect ART, 14 for incorrect fillers, and 3 for both incorrect ART and fillers). We replaced the excluded participants to meet our target sample size of 60 participants. Exclusion criteria were the same as in Experiments 1 and 2.

3.1.2. Procedure

The procedure for Experiment 3 was the same as for the first two except that the Author Recognition task came before the main task instead of after. We shifted the position because we thought this might reduce the number of ART exclusions, which it did compared to Exp. 1 and 2. Our reasoning was that we thought participants were either weary or rushing to finish the ART at the end, especially given that performance on the ART would not cause them to be kicked out of the survey or affect them receiving payment.

3.1.3. Materials and Measures

The third experiment used the same critical stimuli as for the first two experiments, but the audio was cropped in Praat (Boersma, 2001) prior to the pronoun. This allowed us to assess how participants were making predictions given the implicit cause verb and the word *because*.

Our key question was, "Who is likely to be mentioned next?" Participants chose between the names of the two characters.

We used the audio from Experiment 1 for the stimuli, although the pre-pronoun texts were identical across Experiments 1 and 2, so this choice was not critical. All target items were cropped right after *because*. For example, participants might hear *Ana and Liz were attending a karaoke party. Ana idolized Liz because—.* We only used four of the original filler items from Experiment 1 in order to shorten the experiment. Predictions are less constrained than meaning-based interpretations, and we expected that participants would be likely to fatigue if given a longer experiment.

The main task consisted of two practice items, twelve target items, and four filler items. The filler items were the same as in Experiment 1, but we included part of a person's name before the truncation point, e.g., *Will and Ana were at the gym. Will ran a mile with A—.* Thus, participants heard part of the next mentioned character, but not the full name. Participants were instructed to choose the person who was partially mentioned as the person to be mentioned next. Just as in the first two experiments, the 12 target items appeared in two versions: one in which the second sentence contained a subject-biased verb and one in which the second sentence contained an object-biased verb, forming two lists, see **Table 2**. A total of 24 verbs were used in this study, 12 per list (six subject-biased and six object-biased).

Following the audio clip, participants answered two questions. Critical questions always asked, *who is likely to be mentioned next?* The other question was a content question that also functioned as an attention check question for both target and filler items. There were two types of content questions: In the target items, the content question only asked about one character e.g., in the sentences, *Liz and Ana were working out at the gym. Liz loathed Ana because—* participants saw a content question like, *What did Liz do?* In this example, the answer would be *loathed Ana.* In filler items, the content question always asked about both characters e.g., in, *Will and Ana were at the gym. Will ran a mile with A—*, the content question would ask *What did Will and Ana do?* and the answer would be *ran a mile*.

The demographic questions and Author Recognition Task were identical to Experiments 1 and 2.

3.2. Analytical Approach

Data analysis was the same as for Experiments 1 and 2, except that the dependent measure was whether the participant selected the grammatical subject as the character likely to be mentioned next. It was coded as a binary measure with 1 for a grammatical subject selection and 0 for no grammatical subject selection.

3.3. Results

Participants were more likely to identify the subject as likely to be mentioned next in the subject-biased condition (67.5%, SE = .042) than in the object-biased condition (17.78%, SE = .034). A baseline statistical model, testing only the predictor for verb bias, confirmed that there was a significant main effect of verb bias ($\beta = 2.79$, SE = 0.317, $t = 8.80$, $p < .001$).

The primary question of this experiment was whether individuals' prediction of next mention would vary by their performance on the author recognition task and whether this variation would be correlated with variation in an implicit cause bias, a subject bias, or both. Grand-mean centered ART scores were added to the model, as well as the interaction between ART and verb bias.[2] Results revealed a main effect of Condition ($\beta = 2.844$, SE = 0.314, $t = 9.05$, $p < .001$)showed no main effect of ART ($p = .857$), but there was a significant interaction

[2] The ART slope by items was removed from this model because it would not converge.

between ART and Verb bias (β = .047, SE = 0.022, t = 2.14, p =.036), such that as ART score increased, participants were more likely to follow a semantic bias. Participants who scored higher on the ART showed a greater disparity in grammatical subject prediction than did those who scored lower on the ART. Since grammatical subject response was the dependent variable, this is represented by an increase in the selection of the grammatical subject as ART score increased in the subject bias condition, and a decrease in the selection of the grammatical subject as ART score increased in the object bias condition (**Figure 4**).

Again, participants displayed an overall object bias. When the verb was object-biased, participants predicted the character in the object-position (82.2%) would be mentioned next more often than they predicted the character in the subject-position (67.5%) would be mentioned next when the verb was subject-biased.

3.4. Experiment 3 Discussion

Results from Experiment 3 paralleled findings from Experiments 1 and 2: people followed an implicit causality bias when asked to make judgments about who is likely to be mentioned next. Moreover, print exposure influenced the consistency of referential predictions across individuals. People who read more were more likely to predict an implicit cause in a task that did not involve pronoun interpretation.

These results support the conclusion that the results in Experiments 1 and 2 were driven by the information available up until and including the word *because,* and not inferences from the information after the pronoun. Here we saw that predictions before the pronoun were influenced by both verb bias and print exposure. Likewise, in Experiment 1, we saw that pronoun interpretation was influenced by verb bias and print exposure, even though the post-pronominal information was not informative. Together, they suggest that people form referential predictions by the time they finish reading *because* and that these predictions guide pronoun comprehension.

4. GENERAL DISCUSSION

The current study explored whether there were individual differences in ambiguous pronoun resolution in the context of implicit causality verbs, as well as in predictions of likelihood of mention in the same linguistic context. The results of both Experiment 1 and 2 showed that participants' pronoun resolution varied by their scores on the Author Recognition Task (ART), a proxy for print exposure (see figures 2 and 3). Participants with higher ART scores tended to show a stronger semantic bias in both experiments, measured as a larger difference between the subject-biased verbs and object-biased verbs, compared with participants with lower ART scores. Similarly, the results of Experiment 3 showed that participants' prediction of who is likely to be mentioned next varied by scores on the Author Recognition Task, and it did so in the same pattern as for Experiments 1 and 2 (see figure 4). Participants with higher ART scores demonstrated a stronger semantic bias, and those with lower scores demonstrated a weaker semantic bias.

These findings are important for several reasons. First, they add to the body of research that shows that there are individual differences in pronoun comprehension (Arnold et al. 2018; Daneman & Carpenter, 1980; Francey & Cain, 2015; Langlois & Arnold, 2020), and they contribute to the growing body of literature showing individual differences in pronoun resolution as a function of print exposure (Arnold et al. 2018; Arnold, Castro-Schilo, Zerkle, Rao, 2019; Langlois & Arnold, 2020). Critically, our study extends these findings by showing that individual differences in print exposure also influence pronoun resolution in implicit causality contexts.

Results from Experiment 3 also inform the body of literature that explore the time course of referential processing (e.g., Greene & McKoon, 1995; Järvikivi, et al., 2017; Koornneef & Van Berkum, 2006; Long & De Ley, 2000; McKoon, Greene, & Ratcliff, 1993). Although we did not test time course of processing per se, we showed that predictions congruent with a semantic bias can be made prior to any reference in the form of a pronoun or otherwise. In addition, those predictions varied in line with individual performance on the Author Recognition Task. This raises questions about whether print exposure may also relate to differences in the time course of processing.

Previous work has demonstrated that, on average, people tend to expect that the implicit cause will be mentioned. In a sentence fragment like *John admired Bill because he…* people are likely to complete this sentence in such a way that the pronoun refers to Bill, the implicit cause (Brown & Fish, 1983; Garvey, Caramazza, & Yates, 1976; Garvey & Caramazza, 1974; Kehler, Kertz, Rohde, & Elman, 2008; Stevenson, Crawley, & Kleinman, 1994), and in a complete sentence like *John admired Bill because he is great father*, participants consistently show an expectation for the implicit cause, whether they are probed at different points within the sentence or timed in a self-paced reading task (Garnham, Traxler, Oakhill, & Gernsbacher, 1996; Koornneef & Van Berkum, 2006; Long & De Ley, 2000; McDonald & MacWhinney, 1995). So, the question then is: what is it about print exposure that contributes to the variation we see in how people apply implicit cause cues to predictions and interpretations in pronoun resolution?

We consider two accounts for how print exposure might contribute to pronoun and prediction judgments in implicit causality sentences. The Referential Frequency interpretation is that perhaps adults who have had more print exposure than other adults have encountered more instances of implicit causality scenarios, so they have a more robust dataset from which to learn about the frequency of re-mentioning the implicit cause. Evidence from a text analysis shows that in sentences like the ones in our experiment, speakers tend to re-mention the implicit cause more than the other character (Guan & Arnold, under review). That is, the relative frequency of implicit cause re-mention is high in these types of sentences.[3] Through exposure to this pattern, language users may learn that that the implicit cause is a more likely continuation and learn to focus attention on it before any anaphoric reference is mentioned. This interpretation is consistent with other proposals that as we are exposed to language over a lifetime, we become predisposed to common patterns of reference that in turn influence our expectations (Arnold, 1998; 2001; 2010). These patterns could be learned through either spoken or written language. On this view, print exposure matters because it is one type of language experience, and it adds to one's overall lifetime experience with language. People who read more may receive a higher quantity of input than people who read less, thus increasing their exposure to the typical patterns of reference in language. It may additionally be the case that reading can provide a particularly useful type of language input that is helpful for learning about what patterns of reference are more frequent (for further discussion of this point, see Arnold et al., 2018). On the other hand, individual differences in print exposure may also be correlated with individual differences in other language domains, so we cannot conclude that written language is the only source of individual variation. Nevertheless, the robust effect of print exposure is consistent with the idea that language exposure may help people learn discourse statistics.

Alternatively, the Semantic Inference account might explain our finding by building on the idea that semantic constraints matter for predicting the re-mention of a particular referent.

[3] However, the relative frequency of re-mentioning implicit causes may be restricted to only sentences with two human referents; for further discussion see Guan & Arnold, under review.

For example, Kehler et al.'s (2008) model (see also Kehler and Rohde, 2013; Rohde & Kehler, 2014) suggests that people keep track of the likelihood that each referent will be mentioned and that they do so on the basis of semantic representations. In their model, the semantic constraints of implicit causality sentences increase the probability that the implicit cause will be re-mentioned. For example, if you hear *Ana admired Liz because…* as the listener you are likely to expect the speaker to mention the person who is most likely the cause of the admiration. An open question is whether these semantic biases are generated each time people read a sentence or whether they are learned and stored in memory. While the Kehler/Rohde model does not make this explicit, its reliance on the semantic representation is consistent with the idea that listeners make semantic inferences anew for each situation (see also Hartshorne et al., 2015). Conceivably, print exposure could matter because it is related to one's ability to make semantic inferences on a case-by-case basis. Perhaps high print exposure people are better able to use the information at hand to make inferences about likelihood of mention because they have more opportunities to practice doing this. This is in line with findings showing that language skills are stronger for individuals with greater language exposure (Cunningham & Stanovich, 1991; Mani & Huettig, 2014; Montag & Macdonald, 2015; Stanovich & West, 1989; Wells, Christiansen, Race, Acheson, & MacDonald, 2009). Language comprehension involves the ability to make causal judgments and inferences that facilitate local and global coherence of discourse ideas (Graesser, Millis, & Zwaan, 1997). Research shows that the ability to preserve the links between the influx of new information and relevant older information requires higher-order cognitive processes (Graesser et al., 1997) and that there are individual differences in both print exposure and language ability that predict one's ability to do so (Dai, Chen, Ni, & Xu, 2019; Osana, Lacroix, Tucker, Idan, & Jabbour, 2007).

While here we focused on print exposure, a related question is whether individual differences in the use of implicit causality inferences relate to individual variation in reading skill. Given that the Author Recognition task is correlated with measures of reading skill (Moore & Gordon, 2015), we might expect that reading skill also correlates with pronoun comprehension. Indeed, research suggests that it does. For example, Francey & Cain (2015) showed that children with poor reading comprehension skills were less skilled at resolving pronouns in gender-ambiguous cases than children with better reading comprehension skills. Their stimuli required implicit cause inferences by using *because* as the clause connector, e.g., *Michael handed a thank you note to Adrian, after the party, because he was polite*. Each story was followed by a pronoun comprehension question such as *Who was polite?* They found no effects of skill when the gender cue was unambiguous (i.e., two different-gendered characters), but when the pronoun was ambiguous by gender, performance suffered for children with lesser reading skills. A similar conclusion comes from Long and De Ley (2000). They used a probe task to assess processing of implicit causality verbs in two-clause sentences containing a congruent explanation in the subordinate clause. They found that skilled readers were faster at identifying names congruent with the bias than less skilled readers and were significantly more accurate in their responses to the probes. Results also showed that less skilled readers only showed an effect of implicit causality on pronoun resolution after they had integrated information from both clauses.

Thus, previous studies show that using implicit causality information also correlates with reading skills (see also Daneman & Carpenter, 1980). But critically, these studies tested comprehension in a reading task. It is not surprising that reading skill leads to better performance on a reading task. Thus, previous findings could be interpreted as evidence that better readers are

better able to extract information from the written page. The current results provide a critical extension to previous work by demonstrating that print exposure scores correlate with performance on a spoken language comprehension task. This demonstrates that print exposure affects interpretation of pronouns in a way that is not modality specific. Our work also demonstrates that individual differences in pronoun comprehension are specifically linked to differences in prediction inferences.

There are still several unanswered questions about how print exposure relates to pronoun comprehension and prediction. We have suggested two possible explanations for why individuals differ in their usage of implicit causality for pronoun comprehension, the Frequency account, and the Semantic Inference account. Further work is needed to test these accounts. A potential concern for the Semantic Inference account is that it does not offer an obvious explanation for why people with high print exposure tend to follow the subject bias more consistently in other experiments. Arnold et al. (2018, 2019) found this pattern in structures like *Ana was cleaning up with Liz. She…*, and Langlois and Arnold (2020) found this pattern in sentences using transfer verbs (goal/source verbs). Given that reference to the prior subject is a frequent occurrence for both verb types (Arnold, 1998; 2001; Arnold et al. 2018b), the Frequency account can explain this by suggesting that print exposure strengthens representations about which referential patterns are most likely.

On the other hand, a potential concern for the Frequency account is that print exposure has different effects for different semantic biases. In the three experiments reported here, print exposure increased reliance on a semantic bias, implicit causality. But for transfer verbs, Langlois & Arnold (2020) found that print exposure was unrelated to a different semantic bias, the goal bias. Given that goals do tend to be mentioned more often than sources (Arnold, 2001), one would expect print exposure to correlate with this bias. On the other hand, the goal bias is weak. Perhaps stronger semantic biases are more likely to be learned from observing frequencies in the input, or perhaps individual differences are easier to detect for stronger biases. Further work is needed to understand whether print exposure correlates only with some types of semantic biases and how these biases are learned.

While the precise mechanism behind individual differences needs further exploration, the current study makes a valuable contribution to the field by demonstrating that print exposure correlates with the use of implicit causality for both pronoun interpretation and referential prediction judgments. We have shown that this effect is not limited to reading but also affects spoken language comprehension. Our findings build on the results of previous studies that showed that implicit causality makes one referent more predictable (Fukumura & van Gompel, 2010; Kehler et al., 2008; Kehler & Rohde, 2013; Guan and Arnold, under review), and demonstrate that these judgments themselves are modulated by print exposure. Both our prediction findings (Exp. 3) and our findings from the ambiguous-word experiments (Exp. 1) reveal that these biases emerge from information in the first sentence.

We also looked at whether individuals' pronoun resolution would vary with respect to their socioeconomic status (SES). We wanted to determine whether the observed ART effect could instead be explained by participants' SES. Both socioeconomic status (Hecht & Close, 2002; Hoff, 2003; Peterson & Pennington, 2015; Cheng & Wu, 2017) and the author recognition task have been shown to correlate with measures of reading skill, so as a secondary analysis, SES was included as a possible predictor of individual differences. For each experiment, we ran a model including SES and verb bias as predictors without ART, followed by a final model including ART. Overall we found no evidence that socioeconomic status measures explained the

observed correlation between the Author Recognition Task (ART) and pronoun comprehension (Exps. 1 and 2) or the observed correlation between ART and prediction (Exp. 3). A technical report with the full analysis is in preparation.

Our study contributes to our understanding of how language experience relates to language processing. This study is the first to show that print exposure correlates with implicit cause biases at an individual level. This joins a growing set of findings about how referential processing is influenced by individual differences, which together show that print exposure changes the way language is processed at the discourse level.

ACKNOWLEDGEMENTS
We are grateful to Sophia Barsanti, Braxton Hawkins, Darith Klibanow, Heather Mayo, Gustavo Nativio, Lamar Richards, Taylor Steele, and Avery Wall for their assistance in preparing the stimuli and Qualtrics scripts. We are also thankful to artist Jovy Chan (University of Hong Kong) for the creation of the images of our story characters. Data from Experiments 1 and 2 have been adapted from the corresponding author's Master's Thesis, which is cited in references (Williams, 2020).

AUTHOR CONTRIBUTIONS
EW: Conceptualization, Investigation, Data collection, Writing - original draft, Formal analysis.
JA: Conceptualization, Methodology, Writing - review & editing, Supervision, Funding.

CONFLICT OF INTEREST
This research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

FUNDING
This work was funded by NSF grant 1651000 to J. Arnold.

REFERENCES

Acheson, D. J., Wells, J. B., & MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods*, *40*, 278–289. http://dx.doi.org/10.3758/brm.40.1.278.

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, 3*, 137-162. https://doi.org/10.1207/S15326950DP3102_02

Arnold, J. E. (2010). How speakers refer: The role of accessibility. *Language and Linguistics Compass, 4*(4), 187-203. https://doi.org/10.1111/j.1749-818X.2010.00193.x

Arnold, J. E., Castro-Schilo, L., Zerkle, S., & Rao, L. (2019). Print exposure predicts pronoun comprehension strategies in children. *Journal of Child Language*, *46*(5), 863–893. https://doi-org/10.1017/S0305000919000102

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

Au, T. K. F. (1986). A verb is worth a thousand words: The causes and consequences of interpersonal events implicit in language. *Journal of Memory and Language*, *25*, 104-122. https://doi.org/10.1016/0749-596X(86)90024-0

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

Bock, K., & Loebell, H. (1990). Framing sentences. Cognition, 35(1), 1–39. https://doi-org.libproxy.lib.unc.edu/10.1016/0010-0277(90)90035-I

Boersma, Paul (2001). Praat, a system for doing phonetics by computer. Glot International 5:9/10, 341-345.

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*(3), 468–481. https://doi.org/10.1037/0278-7393.31.3.468

Branigan, H. P., Pickering, M. J., Stewart, A. J., & McLean, J. F. (2000). Syntactic priming in spoken production: Linguistic and temporal interference. *Memory & Cognition, 28*(8), 1297–1302. https://doi.org/10.3758/BF03211830

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

Brown, R., & Fish, D. (1983). The psychological causality implicit in language. Cognition, *14*(3), 237–273. https://doi-org/10.1016/0010-0277(83)90006-9

Brysbaert, M., New, B. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990 (2009). https://doi.org/10.3758/BRM.41.4.977

Cheng, Y., & Wu, X. (2017). The relationship between SES and reading comprehension in Chinese: A mediation model. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00672

Corrigan, R. (2001). Implicit Causality in Language: Event Participants and their Interactions. *Journal of Language and Social Psychology*, *20*(3), 285–320. https://doi.org/10.1177/0261927X01020003002

Corrigan, R. (2002). The influence of evaluation and potency on perceivers' casual attributions. *European Journal of Social Psychology*, *32*(3), 363–382. https://doi.org/10.1002/ejsp.96

Corrigan, R. (2003). Preschoolers' and Adults' Attributions of Who Causes Interpersonal Events. *Infant and Child Development*, *12*(4), 305–328. https://doi.org/10.1002/icd.291

Cozijn, R., Commandeur, E., Vonk, W., & Noordman, L. G. M. (2011). The time course of the use of implicit causality information in the processing of pronouns: A visual world paradigm study. *Journal of Memory and Language*, *64*(4), 381–403. https://doi.org/10.1016/j.jml.2011.01.001

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*(2), 264–274. https://doi.org/10.1037/0022-0663.83.2.264

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. https://doi.org/10.1037/0012-1649.33.6.934

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. Cognitive Psychology, 42, 317-367.

Dai, H., Chen, L., Ni, C., & Xu, X. (2019). Literature reading modulates pronoun resolution in counterfactual world: Evidence from event-related potentials. Journal of Experimental Psychology: *Learning, Memory, and Cognition*, *45*(5), 904–919. https://doi.org.libproxy.lib.unc.edu/10.1037/xlm0000620

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

Ehrlich, K. (1980). Comprehension of Pronouns. *Quarterly Journal of Experimental Psychology*, *32*(2), 247–255. https://doi.org/10.1080/14640748008401161

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.

Ferreira, V. S., & Bock, K. (2006). The functions of structural priming. *Language and cognitive processes*, *21*(7-8), 1011–1029. https://doi.org/10.1080/016909600824609

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

Francey, G., & Cain, K. (2015). Effect of imagery training on children's comprehension of pronouns. *Journal of Educational Research*, *108*(1), 1–9. https://doi.org/10.1080/00220671.2013.824869

Fukumura, K., & van Gompel, R. P. G. (2010). Choosing anaphoric expressions: Do people take into account likelihood of reference? *Journal of Memory and Language, 62*(1), 52–66. https://doi.org/10.1016/j.jml.2009.09.001

Garnham, A., Traxler, M. J., Oakhill, J., & Gernsbacher, M. A. (1996). The locus of implicit causality effects in comprehension. *Journal of Memory and Language*, *35*, 517–543. https://doi.org/10.1006/jmla.1996.0028

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

Garvey, C., Caramazza, A., & Yates, J. (1976). Factors affecting assignment of pronoun antecedents. *Cognition, 3*, 227-243.

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

Graesser, A. C., & Millis, K. K., & Zwaan, R.A. (1997). Discourse comprehension. *Annual Review of Psychology*, *48*(1), 163. https://doi.org/10.1146/annurev.psych.48.1.163

Greene, S. B., & McKoon, G. (1995). Telling something we can't know: Experimental approaches to verbs exhibiting implicit causality. *Psychological Science*, *6*(5), 262–270. https://doi.org/10.1111/j.1467-9280.1995.tb00509.x

Grolig L. (2020). Shared Storybook Reading and Oral Language Development: A Bioecological Perspective. *Frontiers in psychology*, *11*, 1818. https://doi.org/10.3389/fpsyg.2020.01818

Guan, S., & Arnold, J. E., (under review). The predictability of implicit causes: testing frequency and topicality explanations.

Hartshorne, J. K., & Snedeker, J. (2013). Verb argument structure predicts implicit causality: The advantages of finer-grained semantics. *Language and Cognitive Processes*, *28*(10), 1474–1508. https://doi-org/10.1080/01690965.2012.689305

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

Hecht, S. A., & Close, L. (2002). Emergent literary skills and training time uniquely predict variability in responses to phonemic awareness training in disadvantaged kindergartners. *Journal of Experimental Child Psychology*, 82(2), 93–115. https://doi.org.libproxy.lib.unc.edu/10.1016/S0022-0965(02)00001-2

Hoff, E. (2003). Causes and consequences of SES-related differences in parent-to-child speech. In M. H. Bornstein & R. H. Bradley (Eds.), *Socioeconomic status, parenting, and child development*. (p. 147–160). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

James, A.N., Fraundorf, S., Lee, E., & Watson, D. (2018). Individual differences in syntactic processing: Is there evidence for reader-text interactions? *Journal of Memory and Language, 102*, 155-181

Järvikivi, J., van Gompel, R. P. G., Hyönä, J. (2017). The interplay of implicit causality, structural heuristics, and anaphor type in ambiguous pronoun resolution. *Journal of Psycholinguistic Research*, *46*, 525-550. https://doi.org/10.1007/s10936-016-9451-1

Kehler, A., & Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics*, *39*, 1–37 https://doi.org/10.1515/tl-2013-0001

Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics*, *25*, 1–44. https://doi.org/10.1093/jos/ffm018

Koornneef, A. W., & Van Berkum, J. J. A. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language*, *54*, 445–465. https://doi.org/10.1016/j.jml.2005.12.003

LaFrance, M., Brownell, H., & Hahn, E. (1997). Interpersonal verbs, gender, and implicit causality. *Social Psychology Quarterly, 60*(2), 138-152. http://dx.doi.org/10.2307/2787101

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

Levin, B. (1993). English verb classes and alternations: A preliminary investigation. Chicago, IL: University of Chicago Press.

Long, D. L., & De Ley, L. (2000). Implicit causality and discourse focus: The interaction of text and reader characteristics in pronoun resolution. *Journal of Memory and Language*, *42*, 545–570. https://doi.org/10.1006/jmla.1999.2695

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

Mani, N., & Huettig, F. (2014). Word reading skill predicts anticipation of upcoming spoken language input: A study of children developing proficiency in reading. *Journal of Experimental Child Psychology*, *126*, 264–279. https://doi.org/10.1016/j.jecp.2014.05.004

McDonald, J. L., & MacWhinney, B. J. (1995). The time course of anaphor resolution: Effects of implicit verb causality and gender. *Journal of Memory and Language*, 34, 543-566.

McKoon, G., Greene, S. B., & Ratcliff, R. (1993). Discourse models, pronoun resolution, and the implicit causality of verbs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1–13. https://doi.org/10.1037/0278-7393.19.5.1040

Montag, J. L., & MacDonald, M. C. (2015). Text exposure predicts spoken production of complex sentences in eight and twelve-year-old children and adults. *Journal of Experimental Psychology: General*, *144*, 447–468. https://doi.org/10.1037/xge0000054

Montag, J. L., Jones, M. N., & Smith, L. B. (2015). The Words Children Hear: Picture Books and the Statistics for Language Learning. Psychological science, 26(9), 1489–1496. https://doi.org/10.1177/0956797615594361

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. https://doi.org/10.3758/s13428-014-0534-3

Nappa, R., & Arnold, J. E. (2014). The road to understanding is paved with the speaker's intentions: Cues to the speaker's attention and intentions affect pronoun comprehension. *Cognitive Psychology*, *70*, 58–81. https://doi.org/10.1016/j.cogpsych.2013.12.003

Osana, H. P., Lacroix, G. L., Tucker, B. J., Idan, E., & Jabbour, G. W. (2007). The impact of print exposure quality and inference construction on syllogistic reasoning. *Journal of Educational Psychology*, *99*(4), 888–902. https://doi.org.libproxy.lib.unc.edu/10.1037/0022-0663.99.4.888

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

Pickering, M. J., & Branigan, H. P. (1999). Syntactic priming in language production. Trends in Cognitive Sciences, 3(4), 136–141. https://doi-org.libproxy.lib.unc.edu/10.1016/S1364-6613(99)01293-0

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: a critical review. *Psychological bulletin*, *134*(3), 427–459. https://doi.org/10.1037/0033-2909.134.3.427

Pyykkonen, P., & Järvikivi, J. (2010). Activation and persistence of implicit causality information in spoken language comprehension. *Experimental Psychology*, *57*, 5–16. https://doi.org/10.1027/1618-3169/a000002

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

Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition*, *19*, 1–30. https://doi.org/10.1016/0010-0277(85)90029-0

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, *24*, 402–433. https://doi.org/10.2307/747605

Stevenson, R. J., Crawley, R. A., & Kleinman, D. (1994). Thematic roles, focus and the representation of events. *Language and Cognitive Processes*, *9*, 519–548. https://doi.org/10.1080/01690969408402130

Stewart, A. J., Pickering, M. J., & Sanford, A. J. (2000). The time course of the influence of implicit causality information: Focusing versus integration accounts. *Journal of Memory and Language*, *42*, 423–443. https://doi.org/10.1006/jmla.1999.2691

Thothathiri, M. & Snedeker, J. (2008). Give and take: Syntactic priming during spoken language comprehension. *Cognition*, *108*(1), 51-68. https://doi.org/10.1016/j.cognition.2007.12.012

Tooley, K. M., & Traxler, M. J. (2010). Syntactic priming effects in comprehension: A critical review. Language and Linguistics Compass, 4(10), 925–937. https://doi-org.libproxy.lib.unc.edu/10.1111/j.1749-818X.2010.00249.x

Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language, 35*, 566-585.

Warriner, A.B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*, 1191-1207. https://doi.org/10.3758/s13428-012-0314-x

Weatherford, K., & Arnold, J. E. (in press). Semantic predictability of implicit causality can affect referential form choice. *Cognition*.

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*(2), 250–271. https://doi.org.libproxy.lib.unc.edu/10.1016/j.cogpsych.2008.08.002

Williams, E. (2020). Language experience predicts pronoun comprehension in implicit causality sentences. [master's thesis]. Chapel Hill (NC): University of North Carolina – Chapel Hill

Table 1. One target item appears across two list, differentiated by verb type.

Exp.	List	Condition	Context Sentence	Target Sentence	Critical Question
1	1	Object-biased	Ana and Liz were attending a karaoke party	*Ana envied Liz because she is a dax.*	Who is a dax? (Ana / Liz)
1	2	Subject-biased	Ana and Liz were attending a karaoke party	*Ana dazzled Liz because she is a dax.*	Who is a dax? (Ana / Liz)
2	1	Object-biased	Ana and Liz were attending a karaoke party	*Ana idolized Liz because she is a great singer.*	Who is a great singer? (Ana / Liz)
2	2	Subject-biased	Ana and Liz were attending a karaoke party	*Ana dazzled Liz because she is a great singer.*	Who is a great singer? (Ana / Liz)

Table 2. One target item appears across two list, differentiated by verb type.

List	Condition	Context Sentence	Target Sentence
1	Object-biased	Ana and Liz were attending a karaoke party.	Ana idolized Liz because—
2	Subject-biased	Ana and Liz were attending a karaoke party.	Ana dazzled Liz because—

Figure Captions:

Figure 1. There are four characters used in this experiment, these are the images that correspond to each character. Top left –Ana, Top right –Liz, Bottom left –Matt, Bottom right –Will.

Figure 2. Experiment 1 results: Participants with higher ART scores show a stronger semantic bias than those with lower ART scores. **LEFT** Each dot represents the participant's average rate of selecting the subject in each condition, such that there are two dots shown per participant. **RIGHT** This graph depicts differences scores, such that each participant's average rate of subject selection in the object-biased verb condition was subtracted from their average rate of subject selection in the subject-biased verb condition. This demonstrates the variability of the semantic bias for each participant.

Figure 3. Experiment 2 Results: Participants with higher ART scores show a stronger semantic bias than those with lower ART scores.

Figure 4. Experiment 3 Results: Participants with higher ART scores show a stronger semantic bias than those with lower ART scores.