REFERENTIAL PREDICTABILITY AND TOPICALITY DIVERGE IN IMPLICIT CAUSALITY

Shuang Guan

Abstract

Intuitively, two important discourse properties of referents, referential predictability and topicality, seem to be related: given that a referent is topical, we expect it to be re-mentioned in the following discourse. On the contrary, the relationship between predictability and topicality is highly debated in current theories on pronoun comprehension and pronoun production. Some studies suggest that discourse topics are predictable, even in part defined by being predictable (Arnold, 2010; Givón, 1983; Prince, 1981). Other research assumes that predictability and topicality are distinct and have separate influences on pronoun interpretation and production (Fukumura & Van Gompel, 2010; Rohde & Kehler, 2014). Much of this work is done involving implicit causality, a phenomenon of pronoun interpretation and re-mention biases towards the referent causally implicated in the event (Garvey & Caramazza, 1974; Kehler et al., 2008). The current study uses implicit causality stories to investigate the relatedness of referential predictability and topicality. We ask three main questions. How is topic status identified in implicit causality contexts? Do contextual factors known to affect judgments of predictability also affect judgments of topicality? Do participants judge the referent chosen as topical to also be predictable? We find that referential predictability and topicality are influenced by different factors; topicality is mostly determined by subjecthood, while predictability is affected by subjecthood and contextual factors. Overall, topicality and predictability judgments do not pattern together, which suggests that they are distinct properties, at least for the implicit causality context.

1 Introduction

Consider the sentences below.

(1) a. Matt adored Will because he ______________.

b. Matt amused Will because he ______________.

Knowing that Matt and Will are both commonly male-gendered English names, the third
person pronoun *he* is technically ambiguous. Despite this, when asked who the ambiguous *he* refers to, most English speakers would choose Will for (1-a) but Matt for (1-b). Something about this construction creates a clear bias in pronoun interpretation, and the change of verb completely flips the interpretation bias. This effect is attributed to the causal relationships implicit in the events. The more common mental representation of the event in (1-a) sees Will as part of the cause of Matt adoring him (Will must have done something for Matt to adore him). In contrast, the more common mental representation of the event in (1-b) sees Matt as part of the cause of him amusing Will (Matt must have done something to amuse Will) (Ferstl *et al.*, 2011; Hartshorne *et al.*, 2015; McDonald & MacWhinney, 1995). This phenomenon is known as *implicit causality* (IC).

Implicit causality comprises a collection of biases that arise in constructions like (1-a) and (1-b) above. Catherine Garvey and Alfonso Caramazza first investigated (and coined the term) implicit causality in 1974. Using similar test sentences, they found that some verbs would lead to a pronoun resolution bias (or equivalently, pronoun interpretation bias) for the character in subject position, (also called N1 for “first noun phrase”) and some verbs would lead to a pronoun resolution bias for the character in object position (also called N2 for “second noun phrase”) (Garvey & Caramazza, 1974). Since Garvey and Caramazza, a wealth of research done on implicit causality has found many more effects, including bias for attributing cause to one character (Brown & Fish, 1983), bias for re-mentioning one character in a story continuation, with or without a pronoun given (Kehler *et al.*, 2008; Koornneef & Sanders, 2013; Pyykkönen & Järviqvist, 2009), and slower reading times for sentences where the re-mentioned pronoun is incongruent with the bias of the verb (Koornneef & Sanders, 2013; Stewart *et al.*, 2000).

It has been shown that pronoun resolution bias in comprehension tasks and re-mention bias in continuation tasks correlate at ceiling rates in equivalent sentences and thus may be seen as one holistic re-mention bias (Hartshorne *et al.*, 2015). On the comprehension side, re-mention bias shows up as the pronoun interpretation bias, and on the production side, re-mention bias shows up as the referent continuation bias. Following Hartshorne & Snedeker (2013), we will henceforth use the general term re-mention bias to refer to both effects.

It may be clear that we have been using relative terms like “bias” and “more predictable” rather than absolute terms, to talk about the judgments made in implicit causality contexts. Although it would be neat to be able to say that implicit causality inferences can be rigidly determined by the syntactic structure of the sentence, this is not the case. Firstly, the re-mention bias is truly a bias and not an all-or-nothing “right answer.” Empirically, not all subjects will agree on who was re-mentioned for the same sentence. Hartshorne & Snedeker (2013) conducted an experiment in which they measured whether the subject or object was chosen as the re-mentioned referent for 264 verbs known to have subject or object biases. They found that for subject-biased verbs, 64.3% of the time, participants’ chose the subject as the re-mentioned referent, while for the object-biased verbs, 81.5% of the time, participants’ chose the object. Additionally, the strength of this bias can be influenced by factors such as the coherence structuring of the surrounding clauses, e.g. whether explicit coherence connectives like *because* are present or not. Furthermore, the re-mention bias can be easily overturned by what comes after. If we finish sentence 1a. in
the following way, *Matt adored Will because he had no one else to care for*, most people would then understand the pronoun as referring to Matt instead of Will. Though this reading is incongruent with the usual bias of the verb *adore*, and the sentence may take a little bit longer to comprehend, it is completely coherent and felicitous. Thus, it is useful to think of re-mention bias as a pragmatic inference that holds up or fails depending on context (Hartshorne, 2014).

Implicit causality is a rich area for studying different aspects of mental representations of language. Past research has used IC to understand how children develop causal schemas, whether these schemas are consistent across languages and cultures, and what kinds of gender and cultural stereotypes are reflected in IC biases (Au, 1986; Ferstl et al., 2011; Hartshorne et al., 2013). In another vein, IC is also a useful environment for studying cognitive or discourse properties of referents, like referential predictability and topicality. Re-mention biases can be seen as a result of comprehenders’ and speakers’ expectations for a character to be referred to again, meaning that this character is more predictable as a referent than the other character (Rosa & Arnold, 2017). Topicality is an ill-defined term generally used to describe the degree to which a sentence or discourse is “about” a referent. Intuitively, predictability and topicality should be highly related; given that a referent is topical, we expect it to be continued in the following discourse, but the relationship between predictability and topicality is highly debated. Some studies suggest that discourse topics are predictable, even in part defined by being predictable (Arnold, 2010; Givón, 1983; Prince, 1981). Other research assumes that predictability and topicality are distinct and have separate influences on pronoun interpretation and production (Fukumura & Van Gompel, 2010; Rohde & Kehler, 2014).

In this study, we use stories involving implicit causality to investigate the relatedness of topicality and predictability. We ask three main questions. The first is how topic status is identified in implicit causality contexts; who do participants pick as being topical? The second is whether factors known to affect judgments of predictability also affect judgments of topicality; we test two fixed effects, verb type (N1-biased versus N2-biased) and the presence/absence of *because*. The third question is whether or not they pattern together in the same conditions; do participants judge the referent chosen as topical to also be predictable? We find that for the implicit causality context, referential predictability and topicality are influenced by different factors; topicality is mostly determined by subjecheckout, while predictability is affected by subjecheckout and contextual factors. Furthermore, topicality and predictability judgments do not pattern together.

Section 2 goes into more detail on the two major factors that influence implicit causality judgments, the same predictors manipulated in in our experiment. Section 3 gives background on the two cognitive properties at hand, predictability and topicality. Section 4 summarizes relevant theories relating predictability and topicality. Section 5 discusses the experiment, and Section 6 concludes with implications and future directions.
2 IC Verbs and Coherence Relations

2.1 Implicit Causality Verbs

Notice that (1-a) and (1-b) constitute a minimal pair, with only the verb being varied. Thus, implicit causality is often associated with the verbs that display especially strong IC biases, and the literature commonly uses the term *implicit causality verbs*.

Exactly what falls under the umbrella term, “implicit causality verbs”? Since implicit causality is a pragmatic phenomenon, verbs must qualify for this category through judgments elicited from native speakers. Theoretically, one could measure the re-mention bias of all verbs able to take animate subjects and objects in an IC context. Most would show no bias for N1 or N2. Those that show statistically significant, systematic bias for N1 can then be labeled *subject-biased* (or *N1-biased*) while those that show bias for N2 can then be labeled *object-biased* (or *N2-biased*). Hartshorne & Snedeker (2013) did exactly this for 720 high-frequency verbs and reported their average object bias. The bias was tested by pronoun resolution in a meaningless context and subjects were asked to choose the re-mentioned character as in the following.

(2) Sally frightens Mary because she is a dax.
Who do you think is the dax?
Sally Mary

Though empirical determination is key, we always want to find some theoretical explanation for empirical phenomenon, thus many semantically-based verb taxonomies have been proposed to predict and classify verbs that have IC biases. Unfortunately, the field has yet to find a taxonomy that adequately categorizes verbs that display significant IC biases. Hartshorne & Snedeker (2013) tested the ability of four well-known IC verb taxonomies, (that of Au (1986), Brown & Fish (1983), Rudolph & Försterling (1997), and Semin & Fiedler (1991)), to predict the IC biases of the 720 verbs and found that they inaccurately predicted that most verbs should be significantly biased (when most verbs are not), as well as failed to group together both subject- and object-biased action verbs. Instead, they find that a new taxonomy based on finer-grain semantics, VerbNet classes, is promising.

The concept of implicit causality is especially relevant for verbs that describe interpersonal or psychological events, so most research on implicit causality has focused on two types of psych verbs called Stimulus-Experiencer (SE) verbs and Experiencer-Stimulus (ES) verbs. These are so named for the thematic roles licensed by the verbs. In these verbs, re-mention is understood to follow the stimulus. Hartshorne & Snedeker (2013) confirmed this general trend, but noted that Levin class 30.1, which is a group of experiencer-stimulus verbs, showed no systematic IC re-mention bias. Overall, they found that Levin classes 31.1 (Stimulus-Experiencer) and 31.2 (Experiencer-Stimulus) were the most internally consistent.

At this point, we note two important misunderstandings not to make. Firstly, “implicit causality verbs” does not imply that implicit causality is only modulated by the verb. The
next subsection will discuss in detail the influence of coherence relations. A comprehensive passage by Pickering & Majid (2007) illustrates all the different variables that may influence the implicit causality inference:

How then do people compute the inference of implicit causality? Various components of the verb’s meaning are of course important [including] how enduring the event is, how concrete it is, whether it is telic or not (Semin & Fiedler, 1988; Rudolph & FoÃŸrsterling, 1997), and how negative its connotative meaning is (Semin & Marsman, 1994). In addition, properties of the participants affect implicit causality. Changing the gender (Lafrance, Brownell, & Hahn, 1997), animacy (Corrigan, 1988, 1992), or typicality (Corrigan, 1992; Garvey et al., 1976) of the participants changes the implicit-causality bias, as do contextual factors that affect focus (Majid, Sanford, & Pickering, 2006). Finally, syntactic form is important, with causal attribution differing for active versus passive constructions (Au, 1986; Garvey et al., 1976; Kasof & Lee, 1994). All of these factors affect the construction of the event representation, and it is this event representation that is used to infer the cause or consequence. (Pickering & Majid, 2007)

Secondly, the term does not claim that what informs implicit causality biases comes from the verbs themselves. In fact, an area of ongoing debate in implicit causality is where the inference comes from. The semantic structure account posits that implicit causality is derived from the literal meaning of the verb, i.e. that the verb encodes causal information in the thematic roles it licenses (Brown & Fish, 1983; Hartshorne et al., 2015). In contrast, the world knowledge account posits that implicit causality is a general, non-linguistic cognitive inference that comes from knowledge about who tends to cause certain types of events (Hartshorne & Snedeker, 2013).

2.2 Coherence Relations

The second major factor in implicit causality is the relation between the implicit causality clause Matt amused Will and the following clause. To put it more generally, IC is affected by discourse structure, or how information from different sentences in a discourse relate to each other in different ways. These various relationships are called coherence relations, and researchers have built up a short list of them which seem to govern discourse. Three common ones are EXPLANATION, RESULT, and PARALLELISM; examples of each used in an implicit causality context are given below (Hartshorne, 2014).

(3) a. Matt amused Will because Matt made a lot of jokes. (EXPLANATION)

    b. Matt amused Will so Will laughed. (RESULT)

    c. Matt amused Will, and John delighted Bill. (PARALLELISM)
Studies have shown that re-mention biases systematically follow the coherence relation. If the clause following *Matt amused Will* is an explanation clause, an ambiguous pronoun will preferentially be resolved to the cause-related referent. If the following clause is a result clause, an ambiguous pronoun will preferentially be resolved to the non-cause-related referent. If the following clause is a parallelism clause, an ambiguous pronoun will preferentially be resolved to the entity playing the parallel role in the previous clause. In the construction with an ambiguous pronoun described above, (3-a) is then basically our introductory example of implicit causality, *Matt amused Will because he...*, so some researchers suggest that implicit causality can be thought of as one sub-case of re-mention biases that arise from discourse structure (Hartshorne, 2014; Kehler et al., 2008). The re-mention bias for cause-related referents in implicit causality is modulated by there being an EXPLANATION coherence relation between the two clauses.

The strength of this bias can then be affected by whether or not there is a coherence connective, an explicit word between the two clauses. Note that a connective is not mandatory for a coherence relation to hold; *Matt amused Will. Matt made a lot of jokes.* is still an example of a discourse organized by the second sentence explaining the first one: the coherence connective simply signals a coherence relation. In our examples above, EXPLANATION is marked by *because*, RESULT is marked by *so*, and PARALLELISM is marked by *and*. Many studies have shown that coherence connectives do have the power of constraining the following clause to be a certain coherence relation, which then affects who is re-mentioned and how strong the bias towards them is (Ferstl et al., 2011). Kehler et al. (2008) posit that the presence of *because* in a prompt doesn’t have its own affect on pronoun interpretation, but is only there to restrict the following clause to be an explanation clause. They found evidence for this in an experiment where for prompts with *because*, *(Tony disappointed Courtney because ____)* and prompts with a full stop *(Tony disappointed Courtney. ____)*, the percentage of sentence completions re-mentioning the cause-related referent was the same between them when only completions judged to be explanations were considered in the second prompt type (Kehler et al., 2008).

### 3 Predictability and Topicality

The way that language is organized by the information contained in words and structures that make up sentences is called *information structure* or *information packaging*. Predictability and topicality are two important information-structural properties that have been shown to guide language comprehension and language production (Arnold et al., 2013).

Predictability comes from the underlying principle that discourses tend to be thematically organized, so information that has already been mentioned is likely to be mentioned again. Predictability can act at multiple levels. Specific words can be predictable, like the last word to “the early bird gets the...” and references to information in the discourse can also be predictable, regardless of the *linguistic form*, the way that something is said. We are concerned with *referential predictability*, the likelihood that a referent will be mentioned again, independent of the linguistic form, or the specific words used in the referring expression (Arnold et al., 2013). We follow Arnold and collaborators in measuring refer-
ential predictability by asking for participants’ judgments of who will be mentioned again (Rosa & Arnold, 2017; Zerkle & Arnold, n.d.a,n).

The other critical property investigated in this experiment is topicality. In the literature, topic still does not have a formal, uniform definition. The canonical and casual definition is “what the sentence or discourse is about” but there is no easy mapping from “aboutness” to linguistic structure across languages or even within languages (Gundel & Fretheim, 2004). Researchers have tried to define it directly on syntactic or phonetic structure as the grammatical subject, the first expression in a sentence, or the expression without stress, in a given context, but these have all been unsatisfactory. Note that subject is a privileged position which usually indicates topicality, and is often considered the default sentence topic (Reinhart, 1981).

We refrain from delving further into how to define topicality in this paper, but we note two aspects of topicality that are relevant to our study. First is that researchers have come to agree on a distinction between discourse topic and sentence topic. Discourse topic is what the larger discourse is about, while sentence topic is what a sentence is about and must explicitly be referred to in the sentence (Reinhart, 1981). The distinction is illustrated in the following example adapted from Reinhart (1981).

(4) Dr. Morgan is a careful researcher and a knowledgeable linguist, but his originality leaves something to be desired.

a. It's about Dr. Morgan.

b. It's about Dr. Morgan's scholarly ability.

Given sentence (4), we could ask what this sentence is about and conceptually come up with two answers. (4-a) would be considered a possible answer for sentence topic; on the other hand, (4-b) is not directly found in the sentence, i.e. there is no expression in the sentence which refers to “Dr. Morgan’s scholarly ability,” so it cannot be the sentence topic. But (4-b) does capture the information given in the sentence, and given the right context - perhaps a situation in which the sentence is part of a reference letter someone is writing for Dr. Morgan - it could be the discourse topic in a larger passage.

Second, the standard in theoretical linguistics is to consider discourse or sentence topic as a binary category; an idea either is the topic or is not the topic. Yet another approach, developed by functional linguists, sees topicality as a scalar property that can be assigned in different degrees to all discourse entities based on a hierarchy (Givon, 1983). This view may be required by linguistic phenomena like variations in referential expression, which fall along a hierarchy of specificity, from unstressed pronouns to highly specific expressions like *the man in the blue hat who gave me crackers* (Arnold et al., 2013).

Although topicality is hard to define, in our study, we measure topicality directly from participants’ judgments. As hearers and readers comprehend a text, they must identify what the topic is, and add new information about the topic to their mental discourse model.
For stories, this can be equated to identifying the protagonist(s) of the story, who serve as discourse topics, and oftentimes, sentence topics, too (Francik, 1986; Sharkey, 1989). Thus, we follow Arnold and collaborators in measuring topicality by asking participants to choose the “main character” of the story (Zerkle & Arnold, n.d.b). Since our stories are only 1.5 sentences long, discourse topic can be considered equivalent to the sentence topic of the target fragment, so we henceforth call what we measured simply just “topic” or “topicality.” Our study can be characterized by using a binary measure of topicality. (However, we leave open the possibility that the non-main character is just judged lower on a continuum of topicality.) This choice of measurement and this specific wording was also informed by the fact that other measures or phrasings of topicality don’t lend themselves well to native speaker introspection.

4 Relevant Theories on Topicality and Predictability

Why are these two information-structural properties, predictability and topicality, important, and how are they related? As previously mentioned, when people communicate, they create a mental model of non-linguistic representations of the characters and events in the discourse, then update those representations as new information is processed. Not all representations are equal though; some information seems easier to access, or more activated, than others. This information-structural property has been called accessibility (or prominence, salience, and focus) (Arnold et al., 2013). Accessibility is another ill-defined term in the literature, but for our purposes, it is sufficient to state that some researchers suggest that accessibility is accrued to entities that are topical and/or predictable (Arnold, 2010). Language scholars agree that accessibility influences language comprehension and language production, but one area of intense debate is exactly what properties of referents which contribute to their accessibility affect pronoun comprehension and pronoun production, as well as if pronoun comprehension and pronoun production are even influenced by the same factors (Arnold, 2001; Arnold et al., 2013; Rohde & Kehler, 2014).

On one side of the debate, predictability and topicality are seen as related properties, which affect both pronoun comprehension and pronoun production. Givón (1983), who theorized topicality as a continuum on which every discourse entity could be evaluated, also suggested that topicality is defined by how continuous a discourse entity is, with the underlying assumption from gestalt psychology that what is continuing is more predictable. Discourse topics should then be predictable. Building on this work, Arnold (2001) proposed that when entities have high referential predictability, they become accessible, which increases speakers’ likelihood of pronominalizing them and guides hearers’ interpretation of pronouns. Furthermore, referential predictability comes from multiple sources, including grammatical subjecthood and thematic role. In a set of story-continuation experiments with transfer-of-possession verbs, as in Lisa gave the leftover pie to Brendan... where Brendan has the thematic role of goal, and Lisa has the thematic role of source, Arnold found that there was a re-mention bias favoring goal referents, a subject bias for pronoun production, as well as a bias to use pronouns more for goals than sources. Rosa & Arnold (2017) also found a re-mention bias favoring goal referents in a written, story-continuation task.
and a verbal, story-retelling task, as well as a bias to use pronouns for goal referents more than for source referents, on top of a subject bias. Arnold (2010) further suggests a connection between topicality and predictability. Since topicality may influence accessibility, referential predictability could be one measure of topicality. So on one hand, one view of pronoun comprehension and pronoun production is that they are both affected by referential predictability, topicality, and accessibility, all correlated in the same direction. Semantic factors do affect pronoun production, and predictability could be seen as a defining feature of topicality (Rosa & Arnold, 2017; Rosa, 2015; Zerkle & Arnold, n.d.a).

On the other side of the debate, predictability and topicality are seen as strictly distinct properties which have separate effects on pronoun comprehension and pronoun production, respectively. Stevenson et al. (1994) conducted a set of sentence-continuation experiments using transfer-of-possession verbs, as in John seized the comic from Bill. He ___. For interpreting the pronoun He at the end, they found a thematic role bias favoring the goal, and a grammatical role bias favoring the subject. They concluded that thematic role affects referential predictability, which affects pronoun interpretation, but that subject bias is independent of thematic role bias and does not affect referential predictability. Fukumura & Van Gompel (2010) conducted a set of sentence-continuation experiments with Stimulus-Experiencer and Experiencer-Stimulus verbs, as in Gary scared/feared Anna after the long discussion ended in a row. This was because..., and found that thematic role affected referential predictability, but did not affect whether or not the re-mentioned referent was pronominalized. Instead, there was a first-mentioned/subject bias for pronoun production. They concluded that semantic biases and referential predictability are not related to pronoun production. Building off of this work, Kehler et al. (2008) claim that there is a dissociation between pronoun interpretation and pronoun production, represented by the Bayesian model below.

\[
P(\text{referent}|\text{pronoun}) = \frac{P(\text{pronoun}|\text{referent})P(\text{referent})}{P(\text{pronoun})}
\]

Pronoun interpretation, represented by \( P(\text{referent}|\text{pronoun}) \), the probability that a referent is re-mentioned given a pronoun, is mainly affected by semantic, contextual factors, while pronoun production, represented by \( P(\text{pronoun}|\text{referent}) \), the probability that a pronoun is used given a re-mentioned referent, is mainly affected by grammatical factors, and there is no overlap. This model is surprising because “it violates the intuition that speakers will pronominalize mentions of referents in just those cases in which their comprehenders would be expected to interpret the pronouns to those same referents” but it accurately captures various empirical results disassociating pronoun comprehension and pronoun production (Rohde & Kehler, 2014). Rohde & Kehler (2014) take this model a step further and consider that grammatical subjuncthood is not really the important factor at play in pronoun production, rather, it’s the information-structural status, topicality. In a sentence-continuation experiment with IC stories, like Amanda amazed Brittany. _______ and Brittany was amazed by Amanda. _______, they tested whether more topical referents were pronominalized more. They found that participants produced more pronouns when referring back to a subject of a passive clause (a stronger marker of topicality) than when referring back to a subject of an active clause (a weaker marker of topicality). So, they ar-
gue that topicality, but not predictability, affects pronoun production, while predictability, driven by semantic factors, affects pronoun interpretation.

Figure 1 below represents the complicated web of interactions between contextual factors, referential predictability, topicality, pronoun comprehension, and pronoun production.

![Diagram of contextual factors and pronoun production](attachment:figure1.png)

**Figure 1**: Solid arrows represent generally accepted relationships of influence in that direction, while dashed arrows represent contested relationships. Finally, the dashed equality sign between pronoun comprehension and pronoun interpretation questions whether the two processes are really mirror images of each other.

In sum, researchers agree that overall, information structure affects language comprehension and language production. It has been widely accepted pronouns are used for entities that are accessible or topical (Arnold et al., 2013; Rohde & Kehler, 2014). Most agree that semantically-driven contextual factors affect referential predictability. Most agree that referential predictability guides pronoun comprehension (Arnold, 2010; Fukumura & Van Gompel, 2010; Rohde & Kehler, 2014). What is still up for debate is whether semantic contextual factors affect pronoun production, by way of referential predictability, whether referential predictability and topicality are related, and ultimately, whether pronoun comprehension and pronoun production are driven by the same factors or not.

The following study directly probes referential predictability and topicality to consider two contested relationships: do contextual factors affect topicality, and are referential predictability and topicality correlated? To our knowledge, neither has been tested before in implicit causality contexts.

## 5 Experiment

For this experiment, the three main research questions were

1. Who do subjects judge to be topical in implicit causality stories?
2. Do contextual factors known to affect referential predictability also affect topicality?
3. Do topicality and predictability pattern similarly on referent choice in the same conditions?

We carried out a survey experiment which manipulated verb type (SE versus ES) within subjects, and connective type (because versus null) and target question type (topicality versus predictability) between subjects. Each participant saw an equal number of both verb types in only one connective condition and one target question condition. For example, a participant would see both verb types all without because and only answer the main character question for all items. We remark that the target questions used here are quite novel. Considering that prior work has relied heavily on story completion tasks, which may be problematic, we instead used metalinguistic judgment tasks created by Arnold and collaborators (Rosa & Arnold, 2017; Zerkle & Arnold, n.d.a).

5.1 Participants

In total, 64 subjects who were native speakers of English residing in the United States or the United Kingdom were recruited on Amazon Mechanical Turk for monetary compensation. Another restriction was that they had not previously completed another one of the Arnold Lab’s experiments, due to collection of Author Recognition Task scores (Stanovich & West, 1989). Ages ranged from 24 to 76 years old.

5.2 Materials and Design

The experiment was carried out on Qualtrics. The surveys were made up of stories involving implicit causality. We used four characters in these stories who had simple and common English names: Ana and Liz, who were explicitly described as female, and Will and Matt, who were explicitly described as male. For each target item, each story consisted of one context sentence in the past progressive tense followed by an incomplete sentence containing an implicit causality verb in the simple past tense. The context sentence served to introduce two characters and the setting. For example, in the without-because condition, Liz and Ana were volunteering at the library. Liz offended Ana... There were a total of 12 distinct target items, and each was shown in two versions. One in which the following fragment contained a SE verb, like offend, and one in which the following fragment contained an ES verb, like dislike. Thus, the ES counterpart to the example given above was Liz and Ana were volunteering at the library. Liz disliked Ana...

There were also 6 filler items which followed a similar format. For all filler items, the context sentence was in the past progressive, and the following fragment had a non-psych verb in the simple past tense; however, for variety, 2 filler items introduced two characters and continued talking about both, 2 filler items introduced two characters and continued talking about only one of them (represented in Table 1), and 2 filler items introduced only one character and continued talking about him/her. Finally, there were two practice items of similar format that were not analyzed. The two practice items and 6 filler items were the same across all surveys. Example items are given in Table 1 along with the number of distinct items of that type used in the experiment. The complete list of items is given
in Appendix A. Note that the term “item” will be used in this paper for what is usually referred to as a “stimulus” in experiments, so as to avoid confusion with the meaning of “stimulus” in the implicit causality context.

Table 1: Example Items

<table>
<thead>
<tr>
<th>Item Type</th>
<th>Item</th>
<th>Total n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice</td>
<td>Will and Matt were doing the laundry. Will folded the clothes with Matt...</td>
<td>2</td>
</tr>
<tr>
<td>Target, SE</td>
<td>Matt and Will were working out at the gym. Matt aggravated Will...</td>
<td>12</td>
</tr>
<tr>
<td>Target, ES</td>
<td>Liz and Ana were volunteering at the library. Liz disliked Ana...</td>
<td>12</td>
</tr>
<tr>
<td>Filler</td>
<td>Matt and Will were eating breakfast. Will took out the cereal...</td>
<td>6</td>
</tr>
</tbody>
</table>

As for choice of verbs, 24 psych verbs with known IC biases were used; half were of the Stimulus-Experiencer type (N1-biased), and half were of the Experiencer-Stimulus type (N2-biased). All SE verbs were taken from Levin class 31.1, and all ES verbs were taken from Levin class 31.2 for multiple reasons (Levin, 1993). First, class 31.1 and 31.2 verbs are easily used with animate subjects and objects, which makes for more natural-sounding test items. Second, these two classes contain a substantial number of the transitive psych verbs commonly used in IC research, thus allowing for comparisons against previous studies. Finally, we wanted to keep verbs’ IC bias similar within their verb type. As discussed previously, Hartshorne & Snedeker (2013) showed that unlike other verb classes, including those under different taxonomies, verb classes 31.1 and 31.2 have internally consistent IC biases. The 12 N1-based verbs had object bias values ranging from 0.17 to 0.27, while the 12 N2-biased verbs had object bias values ranging from 0.80 to 0.89, so across all verbs, stimulus bias values ranged from 0.73 to 0.89, based on (Hartshorne & Snedeker, 2013).

For measuring topicality and predictability, respectively, we had two critical post-story questions: Who do you think is the main character of this story? and Think about the rest of the second sentence in this story. Who do you think will be mentioned? Both were given as 2-alternative forced choice questions, and the character chosen was considered to be more topical/predictable. Finally, we created a content question for each item which asked either “who” was in the story or “what” the characters were doing in the story. This question served as a check that participants were engaged and actually processing the stories. Participants consented to the statement, “I understand that if I do not give enough correct answers I will not get paid. The experimenters will check my answers to ensure reasonable accuracy.”

From the participants’ perspective, in one survey, they saw 2 practice items, 6 target items (3 SE, 3 ES), and 6 filler items. For each item, they answered the content question.

---

1 One exception is the verb, tease. Due to its high object bias from Experiment 2 in Hartshorne & Snedeker (2013) (0.80), we categorized it as an N2-biased verb, though it is considered a Stimulus-Experiencer verb in the Levin and VerbNet classifications.
plus one of the two critical questions. The survey flow was designed so that participants saw the 1.5 sentence story on one page, then clicked to go to the next page with the questions. Thus, successful completions meant that participants were paying attention and storing the mental representation of the story in their short term memory to answer the questions. We instituted a 75% threshold of accuracy on the content questions, so if a participant answered more than 1/4 of the content questions (more than 3) incorrectly, the survey would automatically end, and they were not included.

5.3 Data Analysis

Data analysis was carried out on RStudio. Fixed effects included verb type, connective type, and target question type, which were coded with zero-sum contrasts. For addressing the research questions, mixed-effects logistic regression models were created with Subject Chosen (did the participant choose the subject or not) as the dependent variable. Mixed-effects models are an extension of regular linear models which, in addition to fixed effects, account for random effects, too. Our random effects were participants and items. Essentially, we account for how the manipulated variables can influence individual participants differently and individual items differently. We constructed the models to have maximal random effects; participants were allowed to have random intercepts as well as random slopes for verb type, and items were allowed to have random intercepts and random slopes for the effects of verb type, target question type, and connective type, plus interactions, where appropriate. In one case, the model failed to converge. Under the assumption that it was over-parametrized, the model was re-run with one predictor (specified below) removed.

5.4 Results

Figure 2 below reports the referent choices made by participants in response to each target question for stories containing N1-biased verbs and N2-biased verbs. We first focus on the main character question across the two connective conditions. In both connective conditions, we see a clear bias towards the subject regardless of verb type, and thus no bias towards the causally-implicated referent. This is also visually apparent in Figure 3, where the raw response rate has been graphed as a percentage. This observation was supported by a mixed model of Subject Chosen based on main effects of Verb Type and Connective Type (with Connective Type removed as a predictor for the random effect Item). Only the intercept was significantly different from zero (p < 0.001), thus showing subject bias to be the only determining factor for identifying topicality.

<table>
<thead>
<tr>
<th>Target Questions</th>
<th>Without “because”</th>
<th>With “because”</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subject</td>
<td>Object</td>
</tr>
<tr>
<td>Who do you think is the main character of this story?</td>
<td>40</td>
<td>15</td>
</tr>
<tr>
<td>N1-biased</td>
<td>41</td>
<td>7</td>
</tr>
<tr>
<td>N2-biased</td>
<td>32</td>
<td>16</td>
</tr>
<tr>
<td>Think about the rest of the second sentence in this story. Who do you think will be mentioned?</td>
<td>35</td>
<td>41</td>
</tr>
<tr>
<td>N1-biased</td>
<td>23</td>
<td>25</td>
</tr>
<tr>
<td>N2-biased</td>
<td>37</td>
<td>38</td>
</tr>
</tbody>
</table>

Figure 2: Data table of referent choices for target questions in without and with connective conditions.
Next, we examine referent choice for the next mention across connective conditions. We replicate the established finding that contextual factors do affect predictability. Across both connective conditions, we notice that verb type is influential within each connective condition, and that there seems to be an interaction between verb type and connective type which increases bias towards the stimulus. A graph is given in Figure 4. This observation was supported by a mixed model of SubjectChosen based on VerbType and
ConnectiveType (nothing removed). Verb type had a significant main effect \((p < 0.001)\) and the interaction between verb type and connective type was marginally significant \((p = 0.051)\). Connective type was not significant \((p = 0.56)\), but the intercept was significant \((p < 0.001)\). Thus, predictability is affected by verb type, and the interaction between verb type and connective type. Clearly, different sets of factors affect topicality and predictability. One unexpected finding is that the model states predictability is also affected by a subject bias. We consider explanations for this in the following discussion.

Finally, we focus on each connective condition separately to look at the influence of which target question is asked. Looking at the left halves of Figures 3 and 4 together, we see that in the without because condition, participants overwhelmingly judge the subject as topical in the main character question, but this judgment for predictability is not replicated in the distribution of the next mention question; it’s not true that in this condition, if a referent is considered topical, it is also considered predictable. A mixed effects model of subject chosen based on verb type and target question type (nothing removed) supports this observation. Other than a significant intercept \((p < 0.001)\) the only other significant predictor was target question type \((p < 0.011)\). Thus, on top of a general subject bias, referent choice differs between the target questions. Next, looking at the right halves of Figures 3 and 4 together, the with because condition, we see a complete flip in referent choice for N2-biased items on the next mention question. This suggests that verb type and target question type have a highly influential interaction term, meaning that bias towards the stimulus is only actualized when the target question is the next mention question. A mixed model (nothing removed) shows that all four predictors are significant: the intercept \((p < 0.001)\), verb type \((p < 0.005)\), target question type \((p < 0.025)\), and the interaction term \((p < 0.01)\). Since target question type is an influential main factor for both conditions, referent choices for the main character and next mention question are decidedly different.

### 5.5 Discussion

The results from our experiment suggest that influential factors and resulting referent choices diverges for predictability and topicality. The answer to our first research question, “Who do participants identify as topical?” is that participants overwhelmingly choose the subject. Topicality is only sensitive to subject bias. This is unsurprising given the extensive research showing that subject is a privileged position for the discourse topic. We then also have the answer to our second question, “Do contextual factors known to affect judgments of predictability also affect judgments of topicality?” We have replicated that predictability is sensitive to verb type and the presence of because, but found that topicality is unaffected by both. To answer our third question, “Do participants think that the referents judged topical are also predictable?” we found that eliciting topicality judgments versus predictability judgments affected the referents participants chose. For stories containing N1-biased verbs, the subject bias for topicality and the IC bias of the verb align, so we get roughly equal percentages of subjects chosen as topical and predictable. But for stories containing N2-biased verbs, the subject bias for topicality and the IC bias of the verb (towards the object) point in opposite directions, so we see that the referent choice distributions are negatives of each other. So topicality and predictability do not pattern
together.

One unexpected finding from our experiment is that referential predictability is also affected by a subject bias. For the next mention question, participants had a tendency to choose the subject when the IC bias of the verb aligned with the subject and when it was not aligned with the subject. This is surprising, because some past studies report that subjecehood does not affect predictability in constructions involving transfer-of-possession verbs or IC verbs (Rosa & Arnold, 2017; Fukumura & Van Gompel, 2010). However, a recent study by Zerkle & Arnold (n.d.b) using “with-verb” constructions like Ana is cleaning up with Liz, also finds that there was a subject bias for a next mention question phrased similarly. We posit that this subject bias might be unique to the phrasing of the next mention question. As previously mentioned, the target questions used in our study are a novel measure. Zerkle & Arnold (n.d.a) have noticed that the specific form of the next mention question can affect what biases surface. In an earlier iteration, when the question was phrased as “Who do you think she will talk about next?” they found that it was not specific enough to elicit the expected IC re-mention bias. The issue was amended by adding in an explicit instruction for participants to think about the next part of the story: “Think about the next sentence in this story. Who do you think will be mentioned?” (Zerkle & Arnold, n.d.b). We then adapted this to “Think about the rest of the second sentence in this story. Who do you think will be mentioned?” to fit the 1.5-sentence structure of our experimental items. Thus, it is possible that the subject bias surfacing in our experiment and that of Zerkle & Arnold (n.d.b) is due to the linguistic form of the next mention question, a variable which cannot be compared to past studies that used different measures of predictability.

Overall, our findings are consistent with the disassociated model proposed by Fukumura & Van Gompel (2010) and Rohde & Kehler (2014). If pronoun production is insensitive to contextual factors and only influenced by topicality, then contextual factors should not influence topicality, which is what we found. One explanation for this is that topic is an information-structural status which is usually decided by the speaker before a sentence is spoken, i.e. the speaker has in mind which important entity they want to give information about, so the semantics of the unfolding sentence should not affect the already decided topic. Our findings are incompatible with proposals by Arnold (2010) and Zerkle & Arnold (n.d.a) that predictability could be a defining feature of topicality. If predictability is a defining feature of topicality, then topical referents must be predictable. We do not find that participants think topical referents are also predictable in implicit causality contexts.

We note that the majority of experiments testing the relationships between contextual factors, topicality, predictability, pronoun comprehension, and pronoun production, have been conducted with IC verbs, some with transfer-of-possession verbs. Our experiment showed that referential predictability and topicality diverge for IC verbs, but there is the possibility that these relationships change with different verb types which have different thematic roles or lack strong semantic biases one way or the other. Zerkle & Arnold (n.d.b) found that for the “with-verb” constructions, judgments of topicality and predictability did pattern together; both followed a strong subject bias.
6 Conclusion

In conclusion, this study supports theories that view topicality and predictability as two distinct properties, which then suggest that pronoun comprehension and pronoun production are two processes that happen differently. To paint a picture, when language comprehenders hear or read an implicit causality story with lots of grammatical, semantic, information-structural, and discourse-structural input, the referent that is the subject will be judged the topic, which then (according to Rohde & Kehler (2014) and others) leads them to use pronouns for that referent. Contextual factors don’t affect this pronoun production process. If instead, language comprehenders hear or read an implicit causality story with the same inputs and come across an ambiguous pronoun, they will take into account semantic factors, like which referent is related to the implicit cause of the event or what is the coherence relation between the two clauses, to calculate who is likely to be mentioned again, and then comprehend the pronoun as the congruent referent.

There is much work left to be done. There is still no way to conclusively say which side of the debate is right, no comprehensive and agreed-upon models of reference comprehension and reference production. As previously mentioned, there is the possibility that influential factors could vary for different verb types and for different ways of measuring topicality and predictability. More studies conducted on different verb types with the same measures are needed. Additionally, Rohde & Kehler (2014) point out that studies of pronoun interpretation in psycholinguistics have focused almost exclusively on singular, third person pronouns. But what about plural pronouns? For example, in a story like “Mary gave John a ride to Bill’s house. They...”, there are four different combinations of referents that They could refer to. How could we evaluate the topicality and predictability of combinations of referents? The inanimate pronoun “it” is also a curious case; it can refer to entities, events, entire situations, propositions, descriptions...etc. How could we evaluate the topicality and predictability of these ontological categories? To our knowledge, no existing work tackles these puzzles.

Acknowledgements

I’d like to say a huge thank you to all the faculty who helped me with this thesis. Dr. Arnold, thank you for a wonderful summer research internship that turned into this thesis, and for your remote support since then. Peet, thank you for advising me with the most honest yet encouraging pep talks. Emily, thank you for being my first connection to linguistics at Swarthmore, and your mentorship every since. Professor Grodner, thank you for your critical help with the data analysis. I couldn’t have done it without all of you and the generosity you showed in investing in me.

This work was partially supported by NSF grant 1651000 to J. Arnold, as well as a 2018 Swarthmore Career Services Summer Experiential Fellowship.
References


Appendix A

Table 2: Experimental Items

<table>
<thead>
<tr>
<th>Item ID</th>
<th>Context Sentence</th>
<th>Target Fragment</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Will and Matt were doing the laundry.</td>
<td>Will folded the clothes with Matt (because)...</td>
</tr>
<tr>
<td>P2</td>
<td>Liz and Ana were fixing the bathroom sink.</td>
<td>Liz took out the wrench (because)...</td>
</tr>
<tr>
<td>T1</td>
<td>Will and Matt were fishing with their kids.</td>
<td>Will admired Matt (because)...</td>
</tr>
<tr>
<td>T1</td>
<td>Will and Matt were fishing with their kids.</td>
<td>Will inspired Matt (because)...</td>
</tr>
<tr>
<td>T2</td>
<td>Will and Matt were working at the same company.</td>
<td>Will idolized Matt (because)...</td>
</tr>
<tr>
<td>T2</td>
<td>Will and Matt were working at the same company.</td>
<td>Will fascinated Matt (because)...</td>
</tr>
<tr>
<td>T3</td>
<td>Matt and Will were rooming together in college.</td>
<td>Matt adored Will (because)...</td>
</tr>
<tr>
<td>T3</td>
<td>Matt and Will were rooming together in college.</td>
<td>Matt amused Will (because)...</td>
</tr>
<tr>
<td>T4</td>
<td>Matt and Will were working out at the gym.</td>
<td>Matt loathed Will (because)...</td>
</tr>
<tr>
<td>T4</td>
<td>Matt and Will were working out at the gym.</td>
<td>Matt aggravated Will (because)...</td>
</tr>
<tr>
<td>T5</td>
<td>Liz and Ana were working on a project for class.</td>
<td>Liz despised Ana (because)...</td>
</tr>
<tr>
<td>T5</td>
<td>Liz and Ana were working on a project for class.</td>
<td>Liz bored Ana (because)...</td>
</tr>
<tr>
<td>T6</td>
<td>Liz and Ana were driving to a family reunion.</td>
<td>Liz resented Ana (because)...</td>
</tr>
<tr>
<td>T6</td>
<td>Liz and Ana were driving to a family reunion.</td>
<td>Liz irritated Ana (because)...</td>
</tr>
<tr>
<td>T7</td>
<td>Liz and Ana were volunteering at the library</td>
<td>Liz disliked Ana (because)...</td>
</tr>
<tr>
<td>T7</td>
<td>Liz and Ana were volunteering at the library</td>
<td>Liz offended Ana (because)...</td>
</tr>
<tr>
<td>T8</td>
<td>Liz and Ana were putting up Christmas decorations.</td>
<td>Liz distracted Ana (because)...</td>
</tr>
<tr>
<td>T8</td>
<td>Liz and Ana were putting up Christmas decorations.</td>
<td>Liz teased Ana (because)...</td>
</tr>
<tr>
<td>T9</td>
<td>Will and Matt were attending an office party.</td>
<td>Will dazzled Matt (because)...</td>
</tr>
<tr>
<td>T9</td>
<td>Will and Matt were attending an office party.</td>
<td>Will envied Matt (because)...</td>
</tr>
<tr>
<td>T10</td>
<td>Will and Matt were camping in the woods.</td>
<td>Will frightened Matt (because)...</td>
</tr>
<tr>
<td>T10</td>
<td>Will and Matt were camping in the woods.</td>
<td>Will trusted Matt (because)...</td>
</tr>
<tr>
<td>T11</td>
<td>Ana and Liz were competing in a marathon.</td>
<td>Ana enraged Liz (because)...</td>
</tr>
<tr>
<td>T11</td>
<td>Ana and Liz were competing in a marathon.</td>
<td>Ana hated Liz (because)...</td>
</tr>
<tr>
<td>T12</td>
<td>Ana and Liz were practicing for a ballet performance.</td>
<td>Ana pleased Liz (because)...</td>
</tr>
<tr>
<td>T12</td>
<td>Ana and Liz were practicing for a ballet performance.</td>
<td>Ana worshipped Liz (because)...</td>
</tr>
<tr>
<td>F1</td>
<td>Ana and Liz were watching TV.</td>
<td>Ana changed the channel (because)...</td>
</tr>
<tr>
<td>F2</td>
<td>Ana and Will were changing their furniture.</td>
<td>Ana assembled the bed with Will (because)...</td>
</tr>
<tr>
<td>F3</td>
<td>Matt was playing the piano.</td>
<td>Matt enjoyed himself (because)...</td>
</tr>
<tr>
<td>F4</td>
<td>Liz was preparing a presentation.</td>
<td>Liz created an outline (because)...</td>
</tr>
<tr>
<td>F5</td>
<td>Matt and Liz were going to decorate the house.</td>
<td>Matt went to the store with Liz (because)...</td>
</tr>
<tr>
<td>F6</td>
<td>Matt and Will were eating breakfast.</td>
<td>Will took out the cereal (because)...</td>
</tr>
</tbody>
</table>

*P stands for Practice item, T stands for Target item, and F stands for Filler item.*