EFFECTS OF SEMANTIC ROLE PREDICTABILITY ON REFERENTIAL FORM CHOICE IN EMOTION EVENTS WITH IMPLICIT CAUSALITY

Kathryn C. Weatherford

A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in the Department of Psychology & Neuroscience (Cognitive Program).

Chapel Hill
2019

Approved by:

Jennifer E. Arnold

Peter C. Gordon

Kelly Giovanello

ABSTRACT

Kathryn C. Weatherford: Effects of semantic role predictability on referential form choice in emotion events with implicit causality
(Under the direction of Jennifer E. Arnold)

Referential predictability has been shown to affect comprehension of ambiguous referents. However, the effects of predictability on pronoun production are not as evident. Many models of language production posit that reduced linguistic forms are used when referents are more predictable (e.g. Levy & Jaeger, 2007), yet in several studies using emotion events with implicit causality (e.g. *Jacob impressed Dave because…*), the predictability of the implicit cause had no effect on pronoun use (Fukumura & van Gompel, 2010; Kehler & Rohde, 2013). In contrast, other studies using transfer events (e.g. *Jacob gave the book to Dave…*) have found an effect of goal predictability on pronoun use (Arnold, 2001; Rosa & Arnold, 2017). Why is there a difference across verb types? Has the effect of referential predictability on pronoun use in emotion events with implicit causality been missed previously?

In a novel story retelling paradigm (**Part One**), speakers did use more pronouns to refer to implicit causes versus non-causes, in support of a previously unobserved effect of referential predictability for implicit causes in emotion events. Further, these findings suggest that previous null findings for this event type were possibly due to different methodological requirements across verb types, suggesting avenues for future investigation. **Part Two** directly tested whether the semantic roles for the emotion events used in Part One affect judgments of who will be mentioned next, and found that they do, but that these judgments depend on the availability of the causal coherence relation. Importantly, the increased pronoun use for more predictable implicit causes further supports a role for referential predictability in a model of pronoun production. Finally, **Part Three** investigated production facilitation, through measures of

planning and production, as a mechanism through which predictability might affect referential form choice. While more predictable implicit causes were produced faster, there was no effect of production facilitation on referential form choice.

ACKNOWLEDGEMENTS

This work was partially supported by NSF grant #1348549 awarded to Jennifer Arnold. Thank you to my advisor, Jennifer Arnold, for all of her support, guidance and patience. I also extend heartfelt gratitude to my committee members for their insightful comments and support. Thank you to my cohort members and fellow lab members, including the undergraduate research assistants who aided in collecting and coding the data. Finally, and most especially, thank you to my family. Without your unwavering support in *every* area of my life, you know I would not be here.

TABLE OF CONTENTS

LIST OF TABLES..... viii

LIST OF FIGURES..... ix

INTRODUCTION..... 1

EXPERIMENT 1: Novel story retelling paradigm..... 16

Motivation..... 16

Methods..... 17

Participants..... 17

Materials & design..... 18

Procedure..... 21

Analytic approach..... 23

Results..... 24

Discussion..... 27

EXPERIMENT 2: Predictability ratings..... 28

Motivation..... 28

Methods..... 30

Participants..... 30

Materials & design..... 31

Procedure..... 33

Results..... 33

Discussion..... 35

PART 3: Acoustic analysis of experiment 1 data – speech planning..... 38

Motivation..... 38

Methods..... 43
Participants & Materials..... 43
Procedure..... 43
Results..... 45
Discussion..... 54
GENERAL DISCUSSION..... 56
APPENDIX A: CHARACTER DEPICTIONS..... 62
APPENDIX B: EXPERIMENT 1 STIMULI..... 63
APPENDIX C: EXPERIMENT 2 (A & B) STIMULI..... 66
APPENDIX D: ALL ACOUSTIC MEASURES OF INTEREST..... 68
APPENDIX E: RESULTS FOR OTHER ACOUSTIC MEASURES..... 69
REFERENCES..... 78

LIST OF TABLES

Table 1. Distribution of included items across conditions for Experiment 1..... 23

Table 2. Experiment 1 inferential statistics..... 26

Table 3. Bonferroni corrected estimates for the effect of referential predictability of implicit causality for non-subjects vs. subjects and effect of subject-hood for non-causes vs. causes..... 26

Table 4. Experiment 2a/2b Critical Items only inferential statistics..... 35

Table 5. Estimates for effect of referential predictability of implicit causality in Experiment 2a vs 2b (with connective vs without)..... 35

Table 6. Inferential statistics for target latency to begin speaking the explanation..... 48

Table 7. Correlations between acoustic measures..... 49

Table 8. Pronoun use model output, including target latency as predictor..... 53

Table 9. Bonferroni corrected estimates for the effect of referential predictability of implicit causality for non-subjects vs. subjects and effect of subject-hood for non-causes vs. causes, with target latency as predictor..... 54

LIST OF FIGURES

Figure 1. Example critical trial stimuli for Experiment 1 19

Figure 2. Average pronoun use by grammatical role and semantic role condition 25

Figure 3. Critical trial examples from Experiment 2a & 2b 31

Figure 4. Filler examples from Experiment 2a & 2b 32

Figure 5. Catch filler examples from Experiment 2a & 2b 32

Figure 6. Experiment 2 results for critical implicit causality items 34

Figure 7. Example latency area of interest 39

Figure 8. Planning latency area of primary interest 41

Figure 9. Measures indicative of earlier utterance effects 42

Figure 10. Measures indicative of later utterance effects 42

Figure 11. All acoustic measures 44

Figure 12. Target latency by grammatical role and semantic role 47

Figure 13. Pause duration between target referent and explanation continuation by condition 51

INTRODUCTION

One of the key elements of human language is its variability. Humans are able to both express an idea in a variety of ways, as well as comprehend an idea from a multitude of possible message formulations. This variability allows language to be very flexible, but it can also lead to a great deal of ambiguity. With all the variability possible in both language production and comprehension, how are we so successful at using language? To answer this question, we need to understand the mechanisms that drive this variability. Why is a message produced in one format versus the other? Elucidating the psychological mechanisms behind language variation may also generalize to understanding broader psychological mechanisms.

An area of language use that involves a high degree of variability is that of reference production. Speakers must make choices about how to name and refer to entities in the world, and importantly these references must be in a meaningful form that listeners can understand in order for communication to be successful. Entities may be referred to with reduced forms such as pronouns (e.g. *he, she, it, they*) or more explicit expressions such as names and descriptions. For example, when I lose my keys I may ask "*where are my car keys?", "where are my keys?"* or perhaps an exasperated *"WHERE ARE THEY?"* In addition, I may alter the pronunciation of referents from more acoustically reduced, perhaps even unintelligible, to more acoustically prominent (perhaps with growing frustration as in the third example).

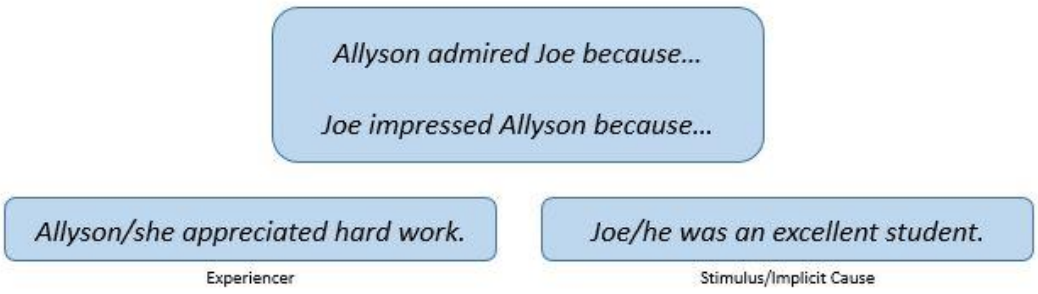
What factors contribute to how referents are ultimately produced? Under what circumstances am I more likely to use *keys* over *they* (or vice versa) and why? One factor may be the likelihood of a referent being mentioned again, or referential predictability. If I have already referenced the keys or stated that I was leaving, subsequent references to the keys may become

more predictable. Does the predictability of references to keys change how I refer to them? This study will address the question of what factors influence referential form by investigating the effects of predictability, and specifically semantic role referential predictability, on referential form choice (i.e. using a name versus a pronoun). This will be done by using specific verbs, such as *admire* and *impress,* that (1) describe emotion events and (2) require arguments that fill the semantic roles of stimulus and experiencer. These verbs are often referred to as implicit causality verbs because one of the arguments, typically the stimulus, is the implied cause of the emotion event felt by the experiencer argument. In both examples below, Allyson is the experiencer and Joe is the stimulus, and implied cause, of what she is experiencing.

Allyson admired Joe because…

Joe impressed Allyson because…

If you have some inclination that the rest of these sentences should include Joe, the likely or implicit cause of the *admiring* or *impressing*, you are not alone. People often expect these continuations to include the implicit cause, and in fact speakers often do continue talking about the implicit cause. In sentences like these, implicit causes are more predictable, or expected, than non-causes. However, this is not an absolute rule, but a question of relative likelihood. It would be just as reasonable if the explanations went on to talk about the experiencer, as in the examples below. The question is, do people use pronouns more when they continue talking about Joe, the more predictable implicit cause, versus Allyson in their explanations?



Predictability is important to language processing

It has long been thought that prediction is a fundamental part of human cognitive processing (Huettig, 2015; von Helmholtz, 1860; James, 1890). Some theorists have even claimed that the human mind is basically a "prediction machine" (Clark, 2013). Many studies find that people simulate and predict the outcomes of the actions of others (Sebanz & Knoblich, 2009), from athletes to ensemble musicians (Keller & Koch, 2008), and that knowing the task of co-actors influences people's own planning and performance (Sebanz, Knoblich & Prinz, 2003). It seems likely that online prediction of others' actions might be an important contributor to both coordinated actions between individuals and to learning. If prediction is fundamental to information processing in general, and language is a large part of the information that humans process, then it is reasonable to assume that prediction also plays a strong role in language processing. The benefits of prediction to language processing may be demonstrated best when language use is seen as a joint activity between interlocutors (Clark, 1996; Pickering & Garrod, 2004). Just as the online prediction of partner's actions has been found to affect the planning and performance of people's own actions in other types of joint activities, prediction may also allow the coordination of planning and task-switching in conversation by facilitating imitation (Prinz, 2006; Pickering & Garrod, 2007). Prediction may therefore constrain the possible variability of language, allowing it to be more effortless and efficient.

The importance of prediction to language use may also be seen in language learning. Chang, Dell & Bock (2006) argue that learning depends on prediction, as individuals make abstractions to new forms based on predictions created from experiences with past forms. This is a possible explanation for how language users can understand and produce new forms that they have never experienced before. For example, Saffran, Aslin & Newport (1996) found that 8 month old babies learned to detect word boundaries in a continuous stream of multisyllabic nonsense words based on statistical dependencies between the syllables. The ability to recognize

and use statistical regularities in language, to learn language, may then depend on prediction skills.

Beyond learning about language and facilitating dialogue, prediction has been found to affect various levels of language in both comprehension and production. In comprehension, prediction makes sense. People are able to predict the actions of others as they unfold, and this can include predicting both *what* others will talk about next and *how* they will talk about it. The goal of the listener is to interpret the speaker's meaning. If they are able to predict (correctly) at least some aspects of the message, comprehension may be faster and easier, memory load may be lessened and the effort to resolve ambiguity and noise may be reduced. Evidence for the effects of predictability on comprehension come from ERP and real-world eye-tracking studies. For example, Kutas & Hillyard (1980) showed participants word-by-word sentences that could end in either a highly predictable semantically appropriate word (based on the context: "He took a sip from the *glass*"), an unexpected and moderately semantically-incongruent word ("He took a sip from the *waterfall*"), an unexpected and strongly semantically-incongruent word ("He took a sip from the *transmitter*"), or an expected semantically appropriate word in an unexpectedly larger font ("He took a sip from the *GLASS*"). They found that the N400, a negative ERP component thought to index semantic processing, was present in both the semantically incongruent conditions, but was attenuated after the more semantically related unexpected word, suggesting that semantic information was pre-activated, or *predicted*, from the context. Similar attenuation was also found for adjectives related, but not unrelated, to a discourse-predictable noun, suggesting that semantic pre-activation occurs throughout the discourse (Van Berkum Brown, Zwitserlood, Kooijman, & Hagoort, 2005). Visual-world eye-tracking studies have also found anticipatory eye movements suggestive of the pre-activation of semantic information for upcoming linguistic information (Altmann & Kamide, 1999; Kaiser & Trueswell, 2004; see Huettig, Rommers & Meyer, 2011 for a review).

The possible effects of prediction on language *production* are less apparent. If a speaker knows and plans the message they want to convey, predictability for what will be said from their perspective is always at ceiling. What benefit would prediction hold for speakers? There is no reason to believe that speakers are making predictions about their own productions. Indeed, Jackendoff (2002) argues against the effects of predictability on language for just this reason. If the goal of language production is to say something meaningful, the predictability of what is produced should play no role. Language production should depend only on the message to be delivered. However, the effects of predictability on production can be seen at various levels of language, from the word level to syntactic ambiguity resolution. Words that are used more frequently, like function words (*a, and, the*), tend to be short whereas less frequent words tend to be longer (*octopus, asparagus*) (Zipf, 1936, 1949). Words that are more probable in the local context have shorter durations (Bell et al., 2003; Jurafsky, Bell, Gregory & Raymond, 2000), and the more information (inversely related to predictability) a word contributes in its context the longer the word tends to be, regardless of its overall frequency (Fowler & Housum, 1987; Piantadosi, Tily & Gibson, 2011). These observations of the effects of frequency and predictability on acoustic duration have been formalized in probabilistic theories of language processing. For example, the probabilistic reduction hypothesis states that more predictable words are reduced (Jurafsky et al., 2000) and the uniform information density (UID) hypothesis states that a positive relationship exists between the information a word conveys and its length in order to convey information uniformly and efficiently (Levy & Jaeger, 2007). It can be extrapolated from these hypotheses that when a referent is highly predictable and less informative it should be produced in a (lexically or acoustically) reduced form.

The evidence for the role of predictability in language production and the theories that have been formalized around that evidence have thus far centered mainly on the predictability of linguistic forms like words (lexical predictability) or syntactic structures. But there are other forms of predictability, and the one of interest to this study is that of referential predictability,

the likelihood that a particular entity will be referred to in a given context. When a referent is predictable, it is consistent or redundant with the previous context and thus may be easier to produce and more fluent (Gahl, Yao & Johnson, 2012; Piantadosi, Tily & Gibson, 2009; Tily & Piantadosi, 2009). This has been supported by the effect on acoustic prominence of manipulations of referential predictability (Kahn & Arnold, 2012; Lam & Watson, 2010; Watson, Arnold & Tenenhaus, 2010; Kaiser, Li, & Holsinger, 2011). The question is, does referential predictability also affect referential form choice?

Referential form choice is driven by accessibility

It is most commonly thought that referential form choice is related to the availability or accessibility of the referent in the discourse model (Chafe, 1976, Prince, 1981; Ariel, 1990; Gundel et al., 1993). In the givenness hierarchy, different referential forms encode cognitive statuses that are related to how and where referents' representations are mentally accessible (Gundel, Hedberg & Zacharski, 1993; Gundel, 2010). For example, unstressed personal pronouns indicate that a referent is accessible in a listener's focus of attention. Many other proposed referential hierarchies are based on the degree of accessibility of a given referent (Ariel, 1990; Givón, 1983). For example, in Ariel's accessibility marking scale, a language producer will select a referential form that matches or signals the ease of accessibility that the listener should have for the referent, given the speaker's intended meaning. The more accessible or available a referent is in the discourse model the greater their representation's activation is on a conceptual level (Givón, 1983; Ariel, 1990; Chafe, 1976, 1994). It is important to note that this representation is a non-linguistic, not a lexical, representation of the entity within a non-linguistic representation of discourse events (i.e. the characters and their actions). Referents are referred to with more explicit forms when they are less accessible, and with more reduced forms such as pronouns when they are relatively more accessible (Arnold, 2010; Ariel, 1990; Brennan, 1995).

In understanding what factors or mechanisms drive referential form choice, an important question to ask then is what makes a referent more or less accessible? Most attempts at defining accessibility include some discussion of the influence of topicality, or what is commonly thought of as what the sentence or discourse is about, and more specifically which entity the speaker means to tell the addressee about (Gundel, 1974; Reinhart, 1981). Topicality may be defined entirely by what has already been talked about (Kehler & Rohde, 2014; Fukumura & van Gompel, 2010), or may include information about both what has already been discussed and the likelihood of upcoming referents, or their predictability (Givón, 1983; Arnold, 2001).

Topics are pragmatic, information structure concepts, not grammatical ones. However, many models use the grammatical subject as at least a proxy for the topic (van Rij, van Rijn, & Hendriks, 2012; Brennan, 1995). Grammatical subject-hood has been found to be a reliable and strong contributor to accessibility (Brennan, 1995; Brennan, Friedman & Pollard, 1987). When the subject of one sentence is subsequently re-mentioned in a following sentence, they are far more likely to be referred to by a pronoun (as in 1A). Conversely, a pronoun is more likely to be interpreted as referring to the previously mentioned subject (Stevenson & Nelson, 1995). This is also true for referents that have been mentioned recently (as in 1C) or in a parallel grammatical position (as in 1B) (Stevenson & Nelson, 1995; Arnold, 2010).

1A: Grammatical subject:	*Lillian met Rob for lunch downtown.* ***She*** *was starving!*
1B: Parallel Grammatical role:	*Lillian met Rob for lunch downtown. The waiter had already sat* ***him*** *at the table.*
1C: Repeated/Recent Mention:	*Lillian and Rob always ate at the same Mexican restaurant. Lillian went there for lunch yesterday. It was a tiny Mexican restaurant and the wait-staff all knew* ***her****.*

However, it should be noted that each of these contributors are likely highly interrelated. Subjects tend to be the topics or focus of discourse, and as such they are likely to be mentioned in a discourse more frequently, more recently and in parallel grammatical position.

Does predictability affect pronoun production?

Some argue that strict topicality, specifically as marked by grammatical roles such as subject or determined entirely by what has already been stated, is all that contributes to accessibility and therefore referential form choice (Kehler & Rohde, 2014; Fukumura & van Gompel, 2010). Others argue that topicality also includes information about what is likely to be discussed in the near future (i.e. persistence or predictability) (Givón, 1983; Arnold, 2001). A third option is that topicality and predictability are separate contributors to accessibility. Regardless of which of these (if any) might be true, the primary question here is whether predictability, directly or indirectly through topicality, influences referential form choice.

There is prior evidence that predictability affects pronoun production. Arnold (1998) argues that discourse features related to pronoun use are also associated with likelihood of re-mention. For example, when an entity has been mentioned in grammatical subject position, there is a greater likelihood that it will be mentioned again. This is also true for parallel grammatical role mention, recent mention and for entities in certain semantic roles and syntactically focused positions (Arnold, 1998).

Tily & Piantadosi (2009) have also found evidence that pronouns are used more often when referents are predictable for the comprehender. Using a corpus of newspaper articles, participants were asked to predict what would be mentioned next when they could only see portions of the text that preceded the next noun phrase. They could choose a previous referent to be mentioned again or "something new" for a new referent. They found that writers (i.e. language producers) used pronouns more often when comprehenders' predictions converged on the same correct referent and longer noun phrases when comprehenders tended to guess incorrectly. When more information is needed to convey a message (i.e. when the message is

less predictable), longer more explicit expressions are required (Jaeger, 2006). This supports an addressee-oriented account of referential form choice where the producer takes into account the predictability of the referent for the comprehender (Brennan & Clark, 1996). However, it may also be that the same factors that affect predictability for the comprehender also affect the producer. Regardless, this demonstrates support for the idea that predictability is related to referential form choice.

How might referential predictability affect referential form choice?

There are several theoretical mechanisms by which semantic role predictability might affect referential form choice. For example, semantic role predictability may influence accessibility by increasing attention to more predictable referents, thus increasing their representations in working memory. Under this mechanism, semantic predictability could work either in tandem with topicality to direct attention, or indirectly by influencing topicality and thus, attention.

As previously discussed, most models of referential form choice state that referential form is based on the accessibility of the referent. It is important to consider whether this accessibility is that of the speaker, the listener or both. If the important accessibility is that of the listener, this suggests that semantic predictability may work through an audience design mechanism, whereby the speaker marks referents with the referential form choice that indicates how accessible the referent should be for the listener. This could be to ensure ease of comprehension and successful communication. Additionally, it could serve the purpose of signaling to the listener that the upcoming information matches their expectations, is easier to process, and/or is more coherent. From an information status perspective, we know that more explicit referential forms signal new and difficult information, while pronouns are used for given, and therefore possibly easier information. In this case, speakers may use a name to signal "Get ready! This is going to be different than you expect or harder to understand."

Semantic predictability may more indirectly affect accessibility of referents for the speaker by allowing planning and production to proceed more efficiently. This facilitation of production mechanism may allow more mental resources to be available to represent the referent and the continuation, may decrease the time between clauses increasing the strength of representations and the link between utterances, or may allow a greater scope of planning such that the information in clauses can be represented together strengthening their link.

How does implicit causality affect referential predictability?

Mae impressed Dave because…

Dave admired Mae because…

In each case, who is likely to be mentioned next? Continuations could re-mention either Mae (*she was a talented piano player*) or Dave (*he knew how much practice it took*), both of them (*they played a beautiful duet together*) or neither of them (*the piano piece was difficult*). However, you may have some intuition that Mae is more likely to be mentioned again in a continuation that relays something about her or something she did that was impressive and worthy of admiration. Many sources of information from both the prior discourse and the situational context may contribute to this likelihood of next-mention, or referential predictability. In both examples, Mae is the stimulus and implicit cause of the emotion that Dave is experiencing. Verbs like *impress* and *admire* belong to classes of verbs that describe interpersonal emotion events and require two arguments to fill the semantic roles of stimulus and experiencer, where the stimulus is often the implicit cause of the emotion. People tend to continue talking about the stimulus or implicit cause of emotion events like these (Fukumura & van Gompel, 2010). Thus, these emotion verbs have been found to have implicit continuation or next-mention biases towards the stimulus or implicit cause of the psychological event (Brown & Fish, 1983; Caramazza, Grober, Garvey & Yates, 1977; Garvey & Caramazza, 1974).

Critically, the next-mention bias for these emotion verbs depends on the relations between utterances or propositions about these emotion events, or the coherence relations

(Kehler, 2002; Kehler, Kurtz, Rohde & Elman, 2008). In both our examples, the coherence relation between clauses is an explanation relation as denoted by the causal connective "because." This drives expectations that the continuation will include an explanation of the cause of the emotion, and thus will include the implicit cause (Kehler, 2002). If the coherence relation in our examples was instead a result relation (e.g. denoted by "so" instead of "because"), the expectation would be for the next clause to contain a causal result or consequence of the previous event, and then the next-mentioned entity is more likely "Dave" (e.g. *Dave was impressed by Mae so Dave/he clapped after her performance*). Importantly, the coherence relation is not dependent on the connective between clauses, and often between sentences there is no connective. However, when present, connectives are usually strong determinants of this relationship.

A useful feature of these emotion verbs is that there are two forms. In the case of verbs such as *impressed*, the stimulus is the first-mentioned (N1) character (the grammatical subject), and thus these verbs have a bias towards continuing with the N1 character in explanation continuations. In contrast, with verbs such as *admired*, the stimulus is the second-mentioned (N2) character (the grammatical object), and thus the next-mention bias is towards the N2 character in explanation continuations. Thus, by comparing across these two related verb classes the effects of predictability (or at least semantic role predictability) can be separated from the known effects of strict topicality and grammatical subject-hood. If semantic predictability has an effect on accessibility and referential form choice over and above that of strict topicality, then implicit cause referents should be pronominalized more often than non-cause referents in explanation continuations, regardless of verb type (N1 or N2 bias) and their grammatical role.

We have already discussed how the role of semantic predictability on referential form is debated theoretically. Empirically, the findings to date have been mixed. Arnold (2001) tested this through both a corpus analysis and experimentally, but with a different class of verbs:

transfer verbs. This class of verb requires arguments that fill the semantic roles of goal and source of the action, but similarly to emotion verbs, there are also two classes. For example, *receive* and *get* require that the subject is the goal of the action, while *give* and *send* require that the subject is the source of the action. In the corpus analysis, Arnold (2001) found that speakers referred more to the subject than the prepositional object and the goal more than the source, supporting a model where both grammatical role and semantic role contribute to referential predictability. In a story continuation experiment, she found that pronouns were used more for subjects than non-subjects, and that for non-subjects, pronouns were used more for goals. This was further supported by empirical work using story continuations also done with transfer verbs, by Rosa & Arnold (2017). Here participants were described a transfer event between two characters (e.g. *The duchess gave the duke a painting)* and then asked to talk about what happened next in the story. Across a series of methods (written, spoken and with/without accompanying pictures), it was again found that goals were pronominalized more often, regardless of grammatical role. Additionally, this study also included an empirical measure of the next-mention predictability of its items, where some participants were simply asked which of the two characters would be mentioned next in the story following the transfer event. Goals were in fact found to be more predictable than sources, while no differences were found between subjects and non-subjects, supporting the idea that semantic role predictability, in addition to strict topicality, does affect referential form choice.

However, the effect of semantic role predictability has not been investigated often with this verb type. In fact, the only other study using transfer verbs did not find an effect of referential predictability, but also did not control for syntactic role (Kehler et al., 2008). Rather, most studies have addressed this question by using implicit causality biases with emotion verbs, and have not found an effect of semantic role predictability on referential form choice, over and above that of strict topicality (Kehler & Rohde, 2013; Rohde & Kehler, 2014; Fukumura & van Gompel, 2010; Kaiser et al., 2011). So the evidence for semantic role predictability effects on

referential form choice is not clear, and brings about the question, why might we see the effects of semantic role predictability in discourse about one event type but not the other?

The reasons may be methodological. The effect of semantic predictability has been seen across paradigm variations for transfer events, but has only been previously tested in passage completions for emotion events with implicit causality. These studies have typically used a sentence or story continuation methodology, where a prompt is given (e.g. *Ana feared John because…)* which participants must complete with a plausible continuation, typically in writing (Kehler & Rohde, 2013; Rohde & Kehler, 2014; Fukumura & van Gompel, 2010; Kaiser et al., 2011, Rosa & Arnold, 2017). This method is beneficial in that it allows tight experimental control over the prompt stimuli used. Participants must also rely solely on the linguistic material they are presented with to create their continuations as opposed to their memory for past events (that cannot be controlled for). However, the tradeoff is that these items are lacking in the context that natural language use typically involves. Often, the only information known about the referents is their name and participants must create continuations entirely independently on the spot. The lack of discourse context may mean that producers are also creating additional context along with their continuations, taxing mental resources and possibly introducing a lot of variation between items and participants. This may be particularly relevant for emotion events with implicit causality, which may already be more complex to represent as compared to transfer events given that they are more abstract, require the speaker to take the mental perspective of at least the emotion experiencer, and vary in their telicity.

Passage completions also require a very incremental form of planning that does not extend to many contexts of natural language use, where people generally know what they are going to talk about before they talk about it. In passage completions, participants comprehend the story prompts as they are presented and then must immediately *create* and plan plausible continuations. Therefore, passage completions may require more mental resources than other types of language use and timing of planning and production may differ such that the effects of

semantic role predictability may be limited or difficult to observe. Speakers usually retrieve a non-linguistic representation of concepts and then the words for those concepts are fit into syntactic frames (Dell, 1986; Levelt, 1989). If the scope of the non-linguistic representation can include multiple events, then those events may be activated simultaneously and a relationship may form between them. However, within the incremental planning of passage completion paradigms, the information represented in each clause may not be activated at the same time and may be represented more as two separate events. Because the next-mention bias depends on the relationship between utterances, information about both utterances may need to be available and conceptualized *in relation to each other* before the referential form has been planned in order for referential predictability to affect accessibility and referential form choice (Arnold, 2013).

Therefore, **Part One** sought to resolve the question of whether semantic role predictability, based on implicit causality with emotion events, affects referential form choice through a novel story *retelling* paradigm. Several adaptations to the traditional passage completion paradigm are included to address some of the possible obscuring aspects of the traditional passage completion paradigm and create more favorable conditions for detecting a predictability effect based with implicit causality. The novel paradigm includes tightly controlled context sentences, sentence prompts and explanations, where participants have access to the continuation explanation prior to having to create its linguistic form. This serves several purposes: (1) to encourage relationships between readily available story events, (2) to eliminate the need for mental resources required for *creating* plausible continuations, and (3) to mimic more natural language use. Additionally, a story-telling context, in which all characters have a visual representation and on-going relationships within a given setting, may also serve to create representational support for these abstract events.

While having the contexts, emotion events and explanations of those events pre-determined has the benefit of creating strong experimental control within and between items, as

well as mimicking more natural language production, this novel design makes it imperative to ensure that the results observed are not due to unintentional systematic differences between items in different conditions, and that the implicit causes in our experimental items are significantly more predictable than the non-causes. The latter will be addressed in a predictability experiment in **Part Two**, where we ask people to make a metalinguistic judgment about who is likely to be mentioned next in our experimental stimuli, while the former is addressed in Part Three, discussed next.

Part Three tests the hypothesis that referential predictability affects referential form choice through facilitation of production by assessing (1) whether latencies to begin utterances, or pauses between different portions of utterances, are related to the semantic and grammatical roles of next mentioned referents, and (2) if those latencies predict referential form choice. Utterances that are easier to construct should be planned and produced more efficiently, resulting in shorter latencies between utterances. If semantic role predictability and topicality as marked by grammatical role influence accessibility of referents, then utterances that continue with reference to the more accessible prior implicit causes and/or subjects may be easier to construct. Additional acoustic analyses of both the duration of utterances and latencies between them across conditions will also be performed in order to address the aforementioned concern that unintentional systematic experimental design factors could mimic a referential predictability effect.

EXPERIMENT 1: Novel Story-retelling Paradigm

Motivation

Experiment 1 was conducted in order to answer the question of whether semantic role predictability, when controlling for grammatical role, induces speakers to use more pronouns when referring to implicit causes versus non-causes within a novel story-retelling paradigm that addresses concerns with previous passage completion studies. This new paradigm addresses these concerns such that: (1) the context of each story is both highly controlled and supportive of the explanation, (2) conditions are equal across items, (3) verbs used are highly N1 or N2 biased, (4) explanations are highly controlled and expand the story without being redundant of earlier information in the story, (5) all short-stories involve the same six characters who all live and work together in the same setting, and (6) only participants who used some variation in referential form were included. This novel story-retelling paradigm allows speakers to continue stories based on facts they learn about characters in a story. Since everything takes place with the same characters in the same general story setting, a richer conceptual representation of the events may be promoted. This paradigm also allows for more natural language use than previous paradigms, as speakers know what they were going to talk about before they speak. Explanations can be tightly controlled, while still fitting a natural explanation form (i.e. explanations are often facts) that can be spoken in a more natural way. Additionally, speakers were instructed to speak in a story-telling manner as though they are talking to a friend about these characters. Thus, the communicative goal of the task and audience is somewhat more concrete than in traditional passage completions.

Additionally, previous studies have had some variation in the specific verbs used when investigating next-mention biases based on implicit causality judgments. These judgments are

based on both the meaning of the verb and what is known about the event it describes (Garvey & Caramazza, 1974; Brown & Fish, 1983; Au, 1986; Crinean & Garnham, 2006 *inter alia*) and the coherence relation between utterances about that event (Kehler, 2002; Kehler et al., 2008). Thus, it is important to look at the verbs that are used to represent both N1 and N2 biases to ensure that they represent similar emotion events, are used in similar ways and have high degrees of bias. To that end, this study used verbs that have previously shown strong N1 or N2 biases (of at least 70%) for explanation continuations, taken from the fine-grained classifications of VerbNet, based on Levin's (1993) verb classes 31.1 ("amaze") and 31.2 ("admire") (Hartshorne, O'Donnell & Tenenbaum, 2015; Hartshorne & Snedeker, 2013). Additionally, critical items always include the "because" connective to ensure that an explanation coherence relation was consistent across items.

Methods

Participants

Seventy-two total participants were drawn from the UNC-CH undergraduate participant pool and were given course credit for their participation. Participants all had normal-to-corrected vision, normal hearing and were native English speakers. Additionally, participants had to use some referential form variation across trials, meeting the minimum criterion of at least one pronoun and at least one character description. Thirteen participants were excluded for lack of variation and three were excluded for technical reasons, leaving 56 (42 female; 75% female) total participants in the analysis. Of these, 50 participants chose to report their age and year in school where the average age reported was 19 years old (standard deviation = 1.9, range = 17-39) and the majority of participants were within the first two years of college (78%), with only 1 participant in graduate school. As an additional descriptive measure, scores on the Author Recognition Task, a proxy measure of print language exposure, were also collected. In this task, participants were shown a list of real (62) and fictional (64) author names, and asked to select the ones they knew to be real (Moore & Gordon, 2014; Stanovich & West, 1989). One point is

given for each correct author selected, and one point is detracted for each incorrect answer. The average score was 16.5 (standard deviation = 8.0).

Materials and Design

In the main task, participants learned facts about a specific character, and then used those facts to retell and complete a mini-story about the daily interactions of a group of characters. Each mini-story, or trial, was about a Mansion and some of the six characters (3 males: *the duke, the driver, the butler*; 3 females: *the duchess, the cook, the maid*) who lived and worked there together. Each trial consisted of three parts: (1) two facts about the character to be continued, (2) a context sentence and (3) a sentence prompt. The character facts both involved the same character, where one fact was a plausible continuation of the mini-story, and the other was not. Placement of the plausible fact (right or left of screen) was counter-balanced between participants within lists. The facts were shown side by side simultaneously on the same screen (Screen 1, *Figure 1*), for as long as participants wished to study them. On the following screen (Screen 2, *Figure 1*), participants read the context sentence and the prompt aloud, continuing with the fact they decided fit the continuation best from those previously learned. The fact learning served the dual purposes of (1) mimicking more natural language production such that participants did not have to make up an explanation, and (2) indicating which character was to be referred to in order to ensure equal numbers of items across conditions (as in previous studies e.g. Rosa & Arnold, 2017; Fukumura & van Gompel, 2010).

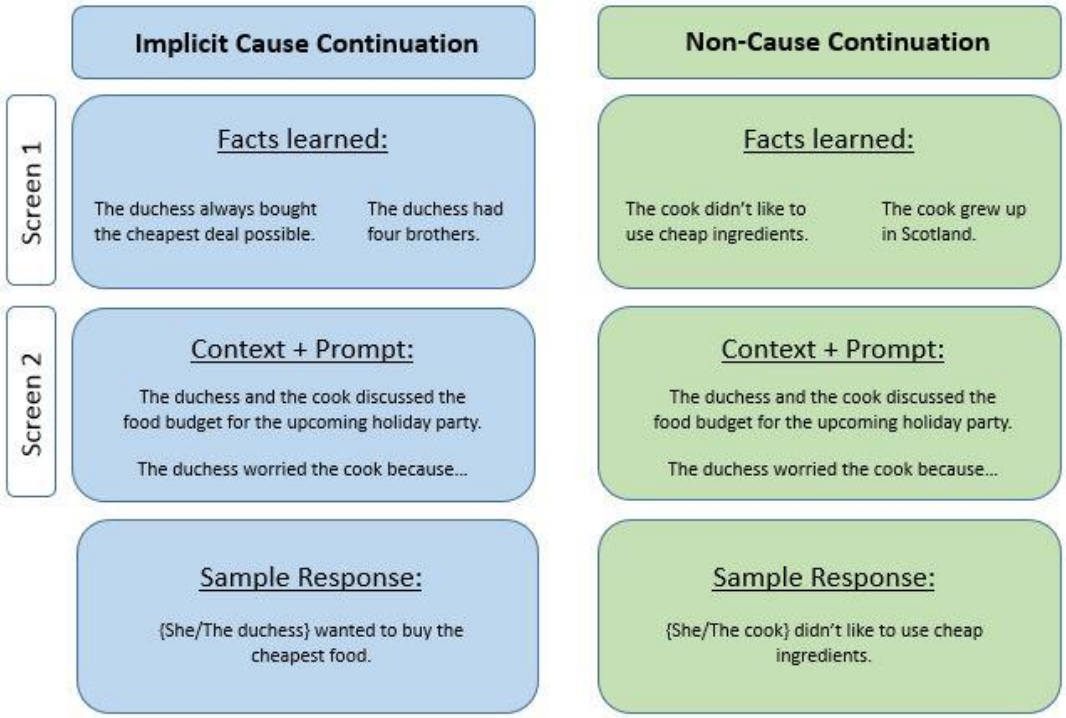


Figure 1: Example critical trial stimuli for Experiment 1: where the implicit cause continuation version would be shown on 1 list, while the non-cause continuation would be shown on the other, with right/left presentation of facts counterbalanced across participants on each list.

There were 24 critical trials and 29 filler trials (*Appendix B*). Filler trials included trials without pronouns in the prompt (13 total), similar in form to the critical trials. These included trials with non-character (5) or double-character continuations (1), or context sentences with more than 1 character (7). Sixteen additional fillers only contained 1 character in the context sentence so that a pronoun was used in the prompt. Thus, slightly less than one-third of total trials included a pronoun within the prompt. Fillers with pronouns in the prompt were used as they are more natural for single subject narratives and they signaled to participants that pronoun use was acceptable in this task.

Critical trials consisted of a context sentence about a shared activity of two same-gendered characters such that there was a compound subject (*Figure 1*). Order of character mention in this compound subject was not controlled for, as it was not thought that this would have an effect on later referential form choice. However, whether the first mentioned character

in the context was also the subject of the prompt was included as a control predictor in the final analyses. Critical trial prompts consisted of *subject character (N1) + implicit causality verb + non-subject character (N2) + "because…".* Character gender and verb type (N1 or N2 bias) were balanced across trials, such that there were 12 trials with male characters and 12 with female characters, and within each of those subsets there were 6 N1 biased verbs and 6 N2 biased verbs. Continuation types were also balanced across verb type and character gender, so that half of all trials with N1 biased verbs continued with the predicted N1 character (*match continuation*) and half continued with the less predicted N2 character (*non-match continuation*). Verb valence was also equal across all trials, genders and continuation type.

The inherent next-mention bias of each verb is not binary, but rather lies on a continuum such that a subject-biased verb does not always illicit a subject re-mention. Thus, a specific verb could show varying degrees of bias depending on the context in which the bias is elicited, and not all subject or object-biased verbs will be biased to the same degree. Several different taxonomies of "implicit causality" emotion or psychological verbs exist (Brown & Fish, 1983; Rudolph & Fosterling, 1997; McKoon et al., 1993; Au, 1986). While all of these taxonomies have similar features and methods for classification, they each assign the bias of verbs slightly differently. For this and other selection reasons, previous studies of this question have used slightly different sets of verbs. We sought to more systematically select verbs such that they would be more consistently biased. Hartshorne & Snedeker (2013) tested the bias of different verbs by asking people who they believed the pronoun referred to in sentences like "*Ann admired Sue because she is a dax."* As compared with more traditionally used taxonomies of "implicit causality" verbs, they found that using the finer-grained verb classes defined by VerbNet (Kipper-Schuler, 2006) predicted verb-biases more accurately and consistently, specifically within the two classes of 31.1 (*frighten, confuse = all N1 bias, except tease = N2 bias)* and 31.2 (*fear, love = all N2 bias)* (Levin, 1993; Lawler, 1993). VerbNet classifies over 270 verb classes based on their shared syntactic alternations, in line with Beth Levin's work showing that

verbs share these alternations because of shared underlying semantic meaning (Levin, 1993). Hartshorne & Snedeker (2013) specifically looked at the biases within these two classes (31.1 and 31.2) because they represent strongly subject (N1) and strongly object biased (N2) verbs, they are readily used with animate subjects and objects, and they include the majority of psychological verbs that have been traditionally used in research on implicit causality, all criteria important for this study. Additionally, as we are specifically interested in the effects of predictability, or next-mention bias, on referential form choice, it would be beneficial to select verbs from the classes that most accurately predict that bias. Therefore, we selected verbs from these two classes within this finer-grained taxonomy based on the strength of the bias found by Hartshorne & Snedeker (2013, exp. 2), with a minimum 70% subject or object bias (Appendix B). The average bias for N1 verbs was 79.08% and the average bias for N2 verbs was 84.5%.

The primary manipulation, whether the explanation character was the implicit cause or non-cause, was manipulated within items and counterbalanced across two lists, such that if a continuation involved the predicted implicit cause character on one list it involved the non-cause character on the other. However, explanation condition was balanced across items and lists, so that all participants saw every item in one condition only. Additionally, each list had two versions across which the presentation order of the facts (left/right of screen) was counterbalanced. Within lists the contextually plausible continuation appeared equally on the left and right of the screen across items.

Procedure

Informed consent was collected from all participants in accordance with UNC-CH IRB protocol. Participants were seated at a Macintosh desktop computer with a 48" flat screen, and fitted with a microphone worn around their neck. Permission to record participants' speech was attained both verbally and in writing, as part of informed consent.

Participants were first given a short description of the task and a general summary of the instructions by an experimenter. They were told that they would learn facts about a group of

characters who all lived together in a mansion. Participants would need to use these facts they learned to help tell short stories about these characters. Critically, the instructions emphasized that they were to speak in a natural story-telling manner, as though they were telling a story to a friend, and that continuations did not need to be the learned facts verbatim but should include the gist.

Following the methods of Rosa & Arnold (2017), participants were then shown a short narrated PowerPoint presentation that introduced the six characters in the story, their illustrated picture, how they were related and the task to be performed. These characters included and were always referred to as: the duke, the duchess, the cook, the driver, the maid and the butler (*Appendix A*). The duke and duchess were married, as were the maid and the butler. The presentation walked participants through two example trials and concluded with four practice trials. Each example trial offered multiple continuation forms, with one modeling both noun phrase and pronoun use in the continuations. In addition, the practice trials gave the experimenter an opportunity to ensure that participants were following instructions, particularly the use of previously learned facts and using a natural story-telling manner. There were also several opportunities for participants to ask questions in order to clarify the task.

The main task consisted of 24 experimental trials and 29 filler trials, for a total of 53 trials. Participants first saw a screen with two facts about one character on the screen. One fact was not related to the trial context, while the other was a coherent and plausible continuation for the upcoming story. Participants were allowed to spend as much time as necessary to learn the facts. They were instructed that they did not need to remember the facts verbatim, but should be able to restate the main gist of the facts in their own words. Directly following the facts, participants were asked to read aloud a context sentence followed by a sentence prompt. In the critical trials, this prompt included two characters of the same gender. Participants were instructed to continue the prompt with the character fact they had just learned that they believed most naturally continued the story.

Following the main experimental trials, participants completed the Author Recognition Task and a short demographic survey. Both surveys were completed on the computer. Participants were then given a short debriefing, asked if they had any questions, and thanked for their participation.

Analytic Approach

The critical question was whether pronominalization of the first character (i.e. the previous subject) mentioned in the explanation would vary as a function of the character's semantic role. Participants' responses were first transcribed by two trained undergraduate research assistants. Critically, transcriptions were then coded for whether participants used a pronoun (*he/she*) or a descriptive noun phrase (*the duchess, the duke*) to refer to the subject character in the continuation explanation. Participants were excluded from analysis if they did not use at least one pronoun or at least one name to describe the subject character in critical trial continuations, as lack of referential form variation may signal that they were not attending to the discourse context (Zerkle & Arnold, 2016). Items were excluded if the participant did not produce the fact that was contextually coherent or if their response did not include the target character in the subject position of the continuation. Additionally, items were excluded if the participant referred to the referent with an incorrect gender (e.g. referring to the cook as *he*). Of the 1,344 items across the 56 included participants, 322 items were excluded. The remaining 1,022 included items were distributed fairly equally across conditions (*Table 1*).

	Non-cause continuation	**Cause continuation**	**GRAND TOTAL**
Non-subj. continuation	229	269	*498*
Subject continuation	233	291	*524*
GRAND TOTAL	*462*	*560*	***1022***

Table 1. Distribution of included items across conditions.

Because pronoun choice was binary, it was modeled with a logistic regression using SAS *proc glimmix*, with centered predictors in a maximal random effects structure. All predictors were binary and centered by coding them as 0/1 and grand-mean centering. For all models, random intercepts and relevant random slopes were included if the model converged and was positive definite.

A control model was first built that included list, side of computer screen the fact appeared on, participant gender and whether the 1st character mentioned in the coordinate subject of the context sentence was also the subject of the prompt to investigate possible effects on pronoun use. The control model included a random intercept for both participants and items and no random slopes, as they were all estimated to be zero. No control variables were found to be a significant predictor and were thus not included in the main model.

Next, the main model was built with the critical predictors of semantic role (implicit cause or non-cause) and grammatical role (subject or object) of the target referent in the prior prompt and their interaction. Again, random intercepts and slopes were included for participant and item if the model converged and was positive definite. The maximal random effects model failed to converge, thus random effects were removed iteratively until the model converged in the order: random slope for item by semantic role, random slope for item by grammatical role, and random intercept for item. Random slopes for participant by semantic role and participant by grammatical role were removed as they were estimated to be zero. Thus, the final model included the random intercept for participant only.

Results

The critical question was whether people would use a pronoun (*he/she*) or a descriptive noun phrase (e.g. *the maid*) when referring to the next mentioned character, and if this choice was dependent on the character's implicit causality. Model details can be found in *Table 2*. Implicit causality was found to have a statistically significant effect on referential form choice, where implicit causes were pronominalized more than non-causes. Additionally, grammatical

subjects were pronominalized significantly more than non-subjects (*Figure 2*). However, these effects were qualified by a significant interaction between semantic role and grammatical role. Visual examination of *Figure 2* suggests that this interaction is due to a stronger effect of implicit causality for non-subjects than subjects. Probing the interaction with Bonferroni corrected Least Squares Means estimates found that more predictable causes were pronominalized more only within non-subjects, while there was no semantic role difference for subjects (*Table 3*). Additionally, subjects were only pronominalized more for non-causes, while there were no significant differences between grammatical roles for causes.

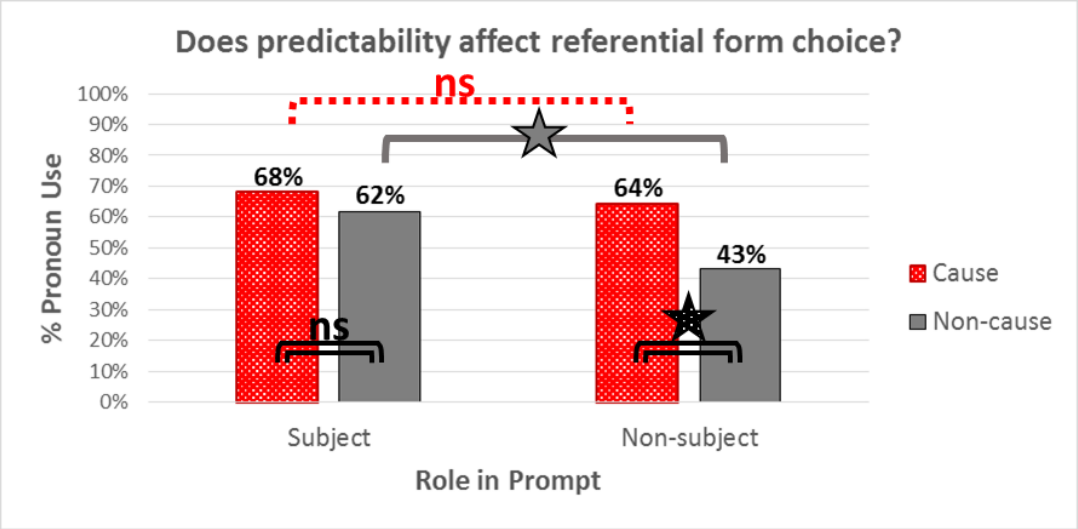


Figure 2. Average pronoun use by grammatical role and semantic role condition

	Effect	Estimate (St.Error)	*t*	*p*
Critical Predictors	Implicit Cause vs. Non-cause	0.65 (0.11)	4.35	<0.0001
	Subject vs. Non-subject	0.60 (0.15)	3.99	<0.0001
Interaction Terms	Thematic role x Grammatical role	-0.82 (0.30)	-2.74	0.006
Control Predictors	Side Fact Shown On (Left/Right)	-----		
	Participant Gender (M/F)	-----		
	List (1/2)	-----		
	1st Context Char. = Subj of Prompt	-----		
Random Effects	Participant	*included*		
	Participant by Cause vs. Non-cause	-----		
	Participant by Subject vs. Non-subject	-----		
	Item	-----		
	Item by Cause vs. Non-cause	-----		
	Item by Subject vs. Non-subject	-----		

Table 2: Experiment 1 inferential statistics

Note. T–values for critical predictors mark their significance. No control variables had *t*-values > 1.5 in the control model, and thus were not included in the final model as indicated by the dashed lines. Excluded random effects are also indicated by the dashed lines.

	Estimate (St. Error)	***t***	***p***	**Adj *p***
Non-subjects: Cause vs. Non-cause	-1.07(0.21)	-5.06	<0.0001**	<0.0001**
Subjects: Cause vs. Non-cause	-0.25(0.21)	-1.20	0.23	0.93
Non-causes: Subject vs. Non-subject	-1.04(.22)	-4.76	<0.0001***	<0.0001***
Causes: Subject vs. Non-subject	-0.23(.20)	-1.11	0.27	1.00

Table 3: Bonferroni corrected estimates for effect of referential predictability of implicit causality for non-subjects vs. subjects and effect of subject-hood for non-causes vs. causes.

Discussion

In a novel story-retelling paradigm, we found evidence that implicit causality does affect referential form choice. Participants used more pronouns when the next-mentioned character was a more predictable implicit cause versus a less predictable non-cause. As expected there was also a significant grammatical role effect such that prior subjects were pronominalized more than prior non-subjects, consistent with the influence of topicality on accessibility. However, these effects were qualified by an interaction between semantic role and grammatical role, such that the semantic predictability effect was only significant for non-subjects, and the grammatical role effect was only significant for non-causes. People showed an increased use of pronouns when referring to either the previous cause or subject, but there was no additional increase in pronoun use for referents that were both causes and subjects. When a referent was a previous subject, a pronoun was used as often when it was a previous non-cause versus a cause. Critically, previous non-subject causes were also as likely to be pronominalized as subject causes, and previous non-subject non-causes, being the least topical *and* the least predictable, were the least likely to be pronominalized. These results are consistent with the view that both topicality *and* predictability, whether directly or indirectly, influence accessibility. However, before further considering the implications of these findings, the predictability of these referents must be confirmed.

EXPERIMENT 2: Predictability Ratings

Motivation

Experiment 2 was conducted primarily as a direct test of whether implicit causality is indeed associated with referential predictability, based on the context sentence and prompts in Experiment 1. This was to verify that implicit causes across items were in fact more expected, or predictable, as the next-mentioned character than the non-causes, supporting the idea that the effects on referential form choice in Experiment 1 were likely due to semantic role predictability. Quantifying the predictability of referents in Experiment 1 for comprehenders can serve as a proxy for the relative predictability of each referent from the speaker's perspective. As previously discussed, the speaker may either model the predictability of referents they expect their audience to have and use that predictability information to guide their referential choices (an audience design account), or the speaker's mental representation of their discourse message may simply be affected by referential predictability similarly to that of the listener's. In either case, the same factors that affect a listener's judgments of referential predictability would also affect predictability for a speaker, and we would expect those judgments to reflect next-mention expectations based on implicit causality. Therefore, participants were asked to choose which of the two referents in each item they believed would be mentioned next, similar to how Rosa & Arnold (2017, Exp.1) measured next mention predictability for transfer events.

Additionally, this experiment allowed a preliminary exploration into possible reasons why predictability effects of implicit causality are seen in this study but not in previous traditional passage completion studies, through an investigation of predictability ratings both before and after the cause coherence relation is marked by the connective "because". Language processing is incremental, and so judgments of predictability may change across an utterance as

new information becomes available. In natural language, speakers hypothetically have access to the information they are planning to communicate earlier than the listener. All of this suggests that differences in the availability of information during planning may change how or if predictability affects referential form choice. One way to investigate this further is by examining how referent predictability changes across an utterance by measuring predictability before and after coherence relation availability (i.e. before and after the connective "because"). Critically, implicit causality next mention bias depends on the coherence relation between utterances (Kehler et al., 2008; Stevenson et al., 1994). In our Experiment 1, speakers have knowledge of the continuations they will use, and may, therefore, be able to form causal coherence relations earlier than in traditional passage completion paradigms. However, the causal relation is not completely apparent, or apparent at all, to comprehenders until the connective "because" in either our novel paradigm or the traditional passage completion paradigm. However, for the producer in traditional passage completions, it is also not apparent until the connective or the creation of the continuation, whichever occurs first. Thus, producers in traditional passage completion paradigms are more like comprehenders, while producers in our novel paradigm are more like real-world speakers. This difference in timing of the availability of continuation and coherence relation information to the producer between paradigms may affect whether referential predictability based on implicit causality can affect referential form choice. If implicit causes are more predictable than non-causes for comprehenders after but not before the connective, this would suggest that implicit causes are not predictable enough early enough in production planning to affect referential form choice in the traditional paradigm. In other words, some knowledge of the causal coherence relation may be necessary for implicit causes to be predictable enough early enough in production planning to affect referential form choice.

If, however, implicit causes are more predictable than non-causes both before and after the connective, this might indicate that there is something else about the traditional passage completion paradigm task that is inhibiting speakers from using referential predictability

information when planning referential form choice. For example, the increased cognitive load of creating a continuation may (1) interfere with the ability to model the listener's discourse representation, (2) include the creation of several possible explanations that might change the predictability of referents in their own mental representations, or (3) might make planning slower and more incremental decreasing coherence between utterances.

Thus, Experiment 2 had two variations, where participants were asked which referent they thought would be mentioned next either with (Experiment 2a) or without (Experiment 2b) the "because" connective. If implicit causes were not more predictable with the connective (i.e. the same exact stimuli used in Experiment 1), the conclusion that the increased pronoun use for implicit causes in Experiment 1 was due to referential predictability would not be supported. In the second version (2b) participants chose the referent they expected to be mentioned next without the "because" connective, in order to compare predictability ratings with the full connective version.

Methods

Participants

123 participants, drawn from Amazon Mechanical Turk, were paid $0.50 for up to 15 minutes of participation. Participants were automatically excluded from completing the experiment unless they were at least 18 years of age, had normal or corrected-to-normal vision, normal hearing, no diagnosed speech disorders, and were native English-speakers (defined as learning English prior to 2 years of age). Participation was also limited to those with IP addresses within the United States and those with HIT approval rates greater than or equal to 95% with at least 10,000 approved HITS. Additionally, participants were also automatically excluded if they responded to more than 2 of 4 catch trials incorrectly. However, this automation failed in the first part of the experiment so 3 participants were excluded after completing the experiment. Therefore, 120 participants were included in the final analyses of the experiment.

Of the 120 participants included for analyses, 61.67% were female (74), 62% were 18-35 years old, 25.6% were 51-80 years old, and 12.4% were 81+ years old. All participants had at least a high school diploma (11.67%), while 25.8% had some college, 14.17% had an associate's degree, 39.17% had a bachelor's degree, and 9.16% had a graduate or professional degree. The average score on the Author Recognition Task was 17.59 (standard deviation = 14.54).

Materials & Design

The 24 critical items from Experiment 1 were divided equally across 3 lists (8 critical items/list), so that there were equal numbers of items with each verb-type and character mentions in both subject and object position. The 3 lists were then duplicated, but with connectives (*because, and then*) removed, for version 2b. Every participant saw only 1 list, but all items were seen across participants. This was done to decrease participant fatigue. Each item consisted of both the context sentence and prompt (*Figure 3*). Participants were then asked to indicate the character they believed would be mentioned next by clicking on the appropriate character description (the same noun phrase descriptions used in the contexts and prompts). Two character choices were provided, which appeared side by side. The order of character choice was counterbalanced across two list sub-versions, so that each character appeared as the left selection on one version of the list and participants only saw one version of the item.

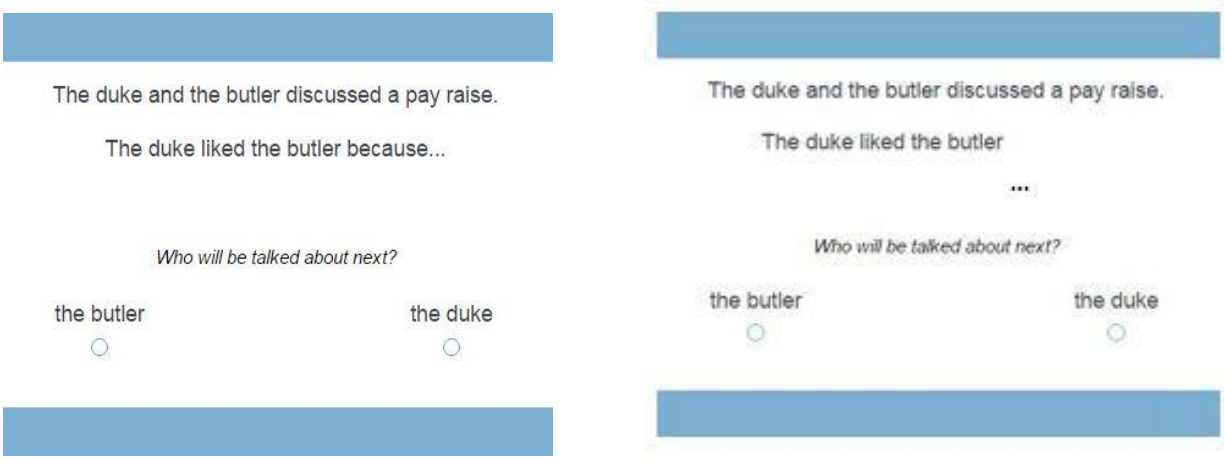


Figure 3: Example critical trial from Experiment 2a (left) & 2b (right)

In addition to 8 critical items, each list contained the same 8 fillers. Four of the fillers were similar in structure to the critical items but involved transfer verbs (goal-source: *got, took*; source-goal: *gave, handed*) (*Figure 4, see also Appendix C*). The remaining four fillers involved prompts with only 1 character as the subject and next mention choices were highly semantically predictable given the context sentence (*Figure 5*). These fillers were considered catch trials, in that participants had to correctly choose the semantically predictable item in at least 2 of the 4 items for inclusion in the study. This was to ensure that participants were reading and comprehending the items and were not randomly selecting answers.

Figure 4: Filler examples for Experiment 2a (left) and Experiment 2b (right)

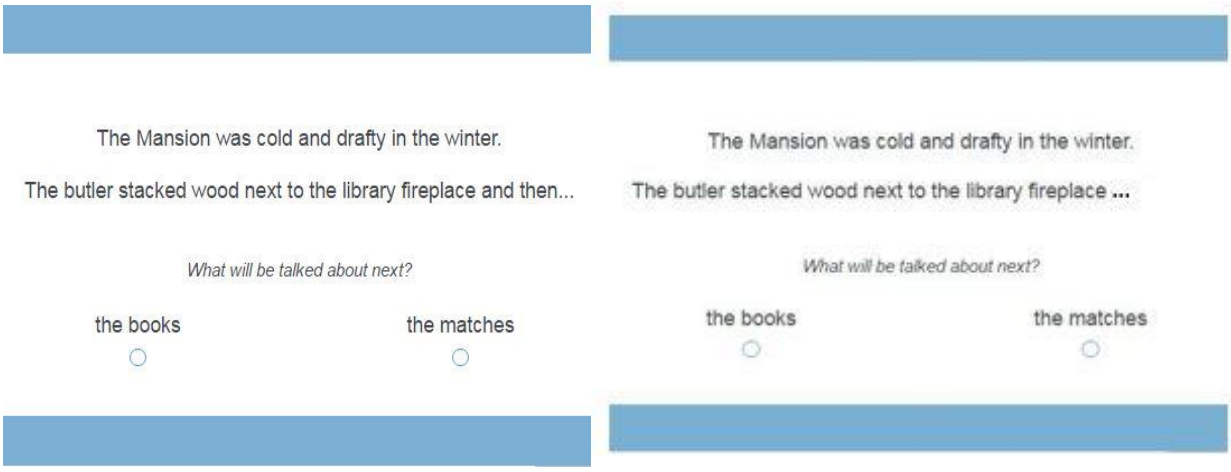


Figure 5: Catch filler example for Experiment 2a (left) and Experiment 2b (right)

Procedure

The study was administered as an online Qualtrics survey through Amazon Mechanical Turk. Participants answered demographic questions followed by the 16 experimental trials and finally the ART. Upon successful completion of all sections, participants were given a unique code in order to receive payment. Each list version was administered to a total of 10 participants so that each stimulus was seen by 20 subjects. All lists were administered on Amazon Mechanical Turk during weekday business hours.

Participants were informed that they would be reading short stories about a group of characters who all lived and worked together in a mansion and the characters were introduced. The participants were then instructed to choose the character that they believed would be mentioned next in each story. Participants were shown two example trials with possible explanations for which character or item might be talked about next in each example. Example trials were of the same structure as filler items, with one involving a character continuation and the other an item continuation.

Results of Experiments 2a/2b

Critical Item data from both experiments were analyzed using the SAS function *proc glimmix* using a single model, where the dependent measure was whether participants selected the character presented on the left as the more likely next-mentioned character, dependent on if the left character was subject/non-subject and implicit cause/non-cause. Because participants were only given two options for the most likely next-mentioned character, and these were counter-balanced so that each character was presented equally on the left or right of the screen, modeling the likelihood of the left character being chosen allowed for both grammatical role and semantic role to be simultaneously modeled as predictors. Model details can be found in *Table 4*. For Experiment 2a, with the coherence relation connective, the next-mention bias for the implicit cause was stronger than in Experiment 2b without the connective. Without the connective, there was a strong bias to select the non-subject regardless of the implicit causality,

while with the connective, there was no grammatical role preference. Overall, there was a main effect of referential predictability based on implicit causality across both experiments, where the cause was selected more often as the next-mentioned character. However, this main effect of implicit causality was modulated by an interaction with the presence of the coherence relation connective "because." Using effect estimates to probe this interaction found that the effect of implicit causality on next mention bias was greater in Experiment 2a, with the connective (β = -1.77 (0.34), t = -5.22, p < 0.0001) (*Table 5*). There was no interaction between grammatical role and thematic role.

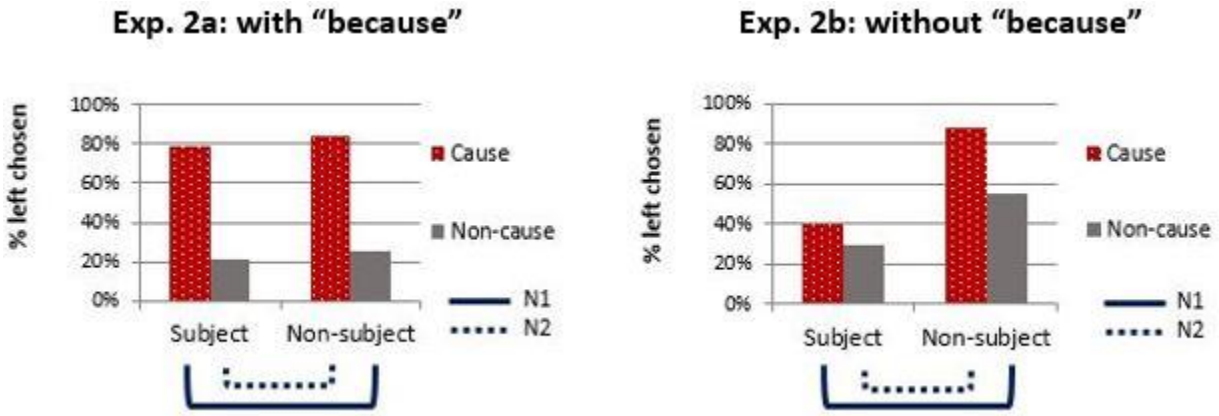


Figure 6: Experiment 2, results of critical implicit causality items only

	Effect	Estimate (St.Error)	*t*	*p*
Critical Predictors	Implicit Cause vs. Non-cause	2.12 (0.17)	12.43	<0.0001**
	Subject vs. Non-subject	-1.2 (0.17)	-6.97	<0.0001**
	"Because" vs No connective	-0.11(0.72)	-0.45	0.66
Interaction Terms	Grammatical role x Thematic role	-0.85	-1.19	0.23
	Grammatical role x Connective	1.51 (0.34)	4.46	<0.0001**
	Thematic role x Connective	1.68(0.34)	4.98	<0.0001**
	Grammatical role x Thematic role x Connective	1.39(0.98)	1.42	0.16
Random Effects	Participant Participant by Cause vs. Non-cause Participant by Subject vs. Non-subject	*included* ----- -----		
	Item Item by Cause vs. Non-cause Item by Subject vs. Non-subject Item by Connective vs. No Connective	included ----- ----- *included*		

Table 4: Experiment 2a/2b Critical Items only inferential statistics, with random slope for connective (i.e. experiment) by items; the model did not converge with other random slopes.

Experiment	**Estimate (St. Error)**	***t***	***p***
2a: with connective	3.04(0.25)	11.98	<0.0001**
2b: without connective	1.27(0.23)	5.53	<0.0001**

Table 5: Estimates for effect of referential predictability of implicit causality in Experiment 2a vs 2b (with connective vs without)

Discussion

Critically, implicit causes were more expected, or predictable, as the next-mentioned character, at least for the events used in this study. This was especially true for causes, regardless of verb-type, in Experiment 2a ("with connective"), supporting the conclusion of Experiment 1 that causes were pronominalized more than non-causes based on greater semantic role predictability. Additionally, this served as a check to demonstrate that the biases of the

verbs used reflect the biases used as selection criteria, reported in Hartshorne & Snedeker (2013, Experiment 2). The average bias found by Hartshorne & Snedeker for the verbs we selected was not significantly different for N1 biased versus N2 biased verbs (although N1 was slightly numerically less on average), and the patterns of predictability in Experiment 2a are congruous with this.

However, it is important to note that the biases found by Hartshorne & Snedeker were based on judgments *after* the causal coherence relation had been signaled by the connective "because." Therefore, while the biases in our Experiment 2a should and do reflect those found in their study, the slightly different pattern of biases seen in Experiment 2b is not inconsistent, but rather orthogonal. This suggests that predictability, at least based on implicit causality, is not consistent across utterances and is significantly affected by the type of continuation that will follow, or more specifically the coherence relation between the event and the continuation, in line with Kehler, et al. (2003).

The fact that the next-mention bias for implicit causes held both before and after the coherence relation was certain suggests that the null findings of traditional passage completion studies is not due to the absence of cause predictability before the coherence relation is known. However, the predictability of causes did increase significantly after the causal connective, leaving open the possibility that causes may just not reach a great enough threshold of expectedness before the coherence relation is known. Additionally, when visually comparing the two graphs in *Figure 6*, it appears that the difference in predictability before and after the connective stems from the fact that causes with N1 biased verbs are not more predictable without the coherence relation, but are with the connective, while N2 biased verbs are cause biased both before and after the connective. Perhaps the traditional passage completion paradigm is really measuring the effects of referential predictability before the coherence relation is known, and thus the null effect on pronoun use is due to not meeting a general predictability threshold either in both verb types or in N1 biased verbs specifically. It is beyond

the scope of the current study, but future research may further investigate what predictability threshold must be met and when in order to affect referential form choice across different verb-types.

Importantly, the results of Experiment 2 validate the conclusions of Experiment 1. Implicit causes are more predictable than non-causes, specifically for the stories used in Experiment 1. Additionally, they suggest that this predictability changes over the course of an utterance. This suggests that planning and production patterns may contribute to the effects of referential predictability on referential form choice. The next section investigates how speech planning and production relates to semantic role, further ensures that there are no unintended systematic differences between conditions, and critically investigates facilitated production as one possible mechanism through which semantic predictability might affect referential form choice.

PART THREE: Acoustic Analysis of Experiment 1 Data – Speech planning

Motivation

The next question to answer is why speakers do in fact use pronouns more when they refer to causes versus non-causes. In this section we report the results of an acoustic analysis testing whether the use of pronouns is the result of production being easier when speakers refer to predictable referents. Additionally, this analysis allowed us to investigate why previous studies with this verb type have not found an effect of semantic predictability, and provided further support for the effects seen in Experiment 1.

First and most critically, this analysis served as an investigation of whether predictability affects referential form choice through facilitated production. As previously discussed, facilitation of production could affect referential form choice in several possible ways. For example, if more predictable events and referents are faster and easier to represent and retrieve and therefore faster and easier to produce, accessibility may increase as the time between message formulation and production decreases, or, as Arnold & Nozari (2018) suggest, hypothetically by increasing the amount of mental resources (i.e. working memory) available to represent referents. Thus, more predictable referents become more accessible and therefore more likely to be pronominalized, albeit not directly due to their predictability. Similarly, with greater available mental resources and a tighter time course of planning, the scope of planning could increase such that there is greater discourse connectivity between events and continuations. This increased connectivity could have an iterative effect in that as the relationship between utterances strengthens, the representations of the individual elements (i.e. the event and the continuation) strengthens, including the accessibility of the referents. Thus, the stronger the discourse coherence the stronger the representations may be, increasing

accessibility. While determining specifically *how* facilitation affects accessibility and referential form choice is beyond the scope of this study, our goal with this analysis is to test whether predictability is related to response planning, by using pausing and word duration as a metric of the time needed to plan. If semantic role predictability does affect referential form choice by facilitating production then we would expect to see less difficulty planning and less time to begin speaking continuations when the next-mentioned character is the more expected implicit cause, and that this more efficient production would lead to greater use of pronouns. Thus, in the examples below (*Figure 7*), we would first expect to see shorter pauses between *worried* (the N1 biased verb with implicit causality) and a continuation about the duchess (the implicit cause) (top example box) versus a continuation about the cook (the non-cause) (bottom example box), and second greater use of pronouns when these pauses are shorter.

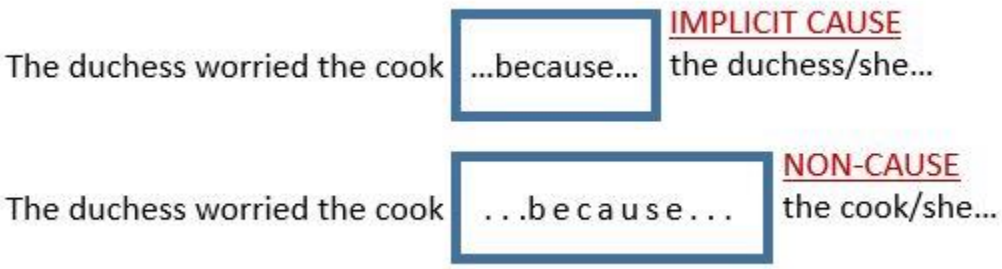


Figure 7. The boxes indicate where production would be expected to be affected by the semantic role predictability of the next-mentioned character.

However, previous evidence using transfer events has not supported the hypothesis that semantic predictability affects accessibility or referential form choice through facilitation of retrieval, planning or production. Rosa & Arnold (2017) found that, while predictable goal continuations were produced faster, this more efficient production was not related to referential form choice. Nevertheless, there is also evidence that emotion events with implicit causality are represented and talked about differently from transfer events. This is seen by comparing the consistency of a predictability effect for transfer events across experimental paradigms to that of

emotion events with implicit causality, where the effect has only been found with this novel paradigm.

Additionally, in this novel story-retelling paradigm participants must produce *both* the emotion event (at least in so far as they are asked to read it out loud) and the continuation, whereas in the traditional passage completion paradigm participants only have to produce the continuation. The traditional paradigm is a turn-taking dialogue and is even somewhat akin to question answering, whereas the story re-telling paradigm is more of a narrative task. Perhaps, when a speaker must plan and produce a series of related utterances in order to convey a story, the timing of planning and production is such that facilitation of production contributes more significantly to accessibility and referential form choice.

This paradigm also allows an investigation of earlier effects on production due to the difficulty of planning the explanation referent. For example, if any planning of the explanation occurs while the prompt is being produced and if non-cause and/or non-subject referents are more difficult to plan, prompt duration or disfluency may be greater. Similarly, if planning and production of a non-cause or non-subject referent is more difficult this might make concurrent planning of the rest of the explanation fact (i.e. down-utterance planning) more difficult or later to occur, which would be indicated by longer pauses after the target referent or longer speech durations and/or more disfluencies within the explanation facts. However, any differences in durations or disfluencies of explanations between continuation conditions would need to be carefully examined, as explanations were created by the experimenter and differ between conditions. Thus, differences in explanation production might also indicate inadvertently systematic differences in the kinds of explanations that were created between conditions. Therefore, effects of continuation condition on both early and later speech production were investigated, where early effects would be suggestive of the ease with which different referent types are produced and later effects would require further investigation to ensure the effects seen in Experiment 1 were not due to stimuli design.

Therefore, the purposes of these analyses were three-fold:

1. *Are pauses before the first character mentioned in the explanation related to the explanation's semantic role condition?*

If the more predictable implicit cause referents were preceded by shorter pauses at re-mention (*Figure 8*), this would indicate that cause continuations are easier/faster to plan, replicating the findings of Rosa & Arnold (2017) for transfer events. Given that Experiment 2 found predictability of causes to increase significantly across the utterance, extending this finding to this event type would also demonstrate that semantic role predictability reaches a great enough threshold early enough in planning to affect at least some aspects of speech production, further supporting the conclusion that the effect seen in Experiment 1 is due to predictability differences for the next-mentioned character in the explanation. If, however, cause continuations were not shown to be faster/easier to plan, this would be direct evidence against the possibility that predictability affects referential form choice through facilitation of production.

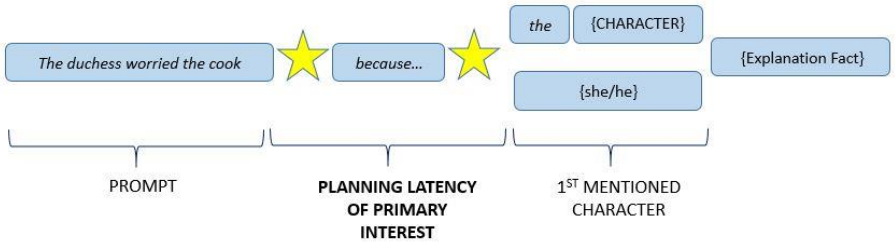


Figure 8. Planning latency of primary interest is bolded and includes the pause between the prompt and *because*, the duration of *because*, and the pause between *because* and the 1st referent mentioned in the continuation.

2. *Are other speech or pause measures related to the explanation's semantic role condition?*

Although the latency period was where we expected to find the greatest effects of response planning, it is also possible that participants began planning earlier in the task. As a set of control analyses, I therefore investigated whether continuation condition affected 3 regions early in the utterance: 1) speech durations and disfluencies of the context sentence; speech

durations and disfluencies of the prompt, and the pause duration between the context and prompt (*Figure 9*).

Additionally, if pauses between the first mentioned character in the continuation and the rest of the explanation differ by continuation condition, this could indicate that production of the target referent affects later production, supporting the idea that predictability does reach a high enough threshold to affect language production. However, if the disfluencies or durations of the rest of the explanation fact (i.e. later utterance measures) also differ by continuation condition, this could also indicate that that there are inadvertent stimuli differences between conditions and further investigation would be required. Therefore, as an additional control analysis, the effect of continuation condition on each of these measures was also examined (*Figure 10*).

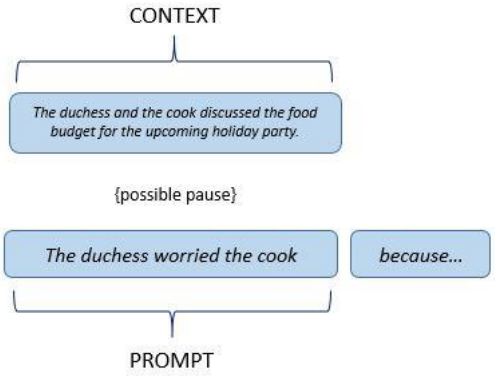


Figure 9. Measures indicative of earlier utterance effects of continuation type include the duration or disfluencies of the context sentence and the prompt, and the pause duration between these utterances.

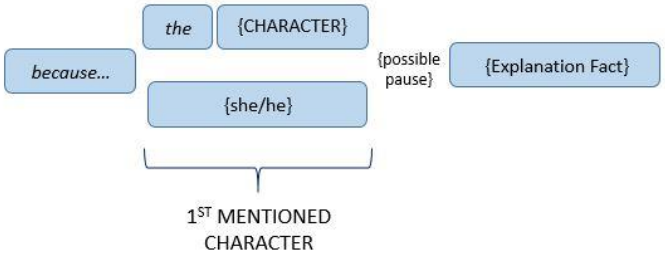


Figure 10. Measures indicative of later utterance effects of continuation type include the duration or disfluencies of the explanation fact and the pause duration between the 1st mentioned character and the rest of the explanation.

3. *Are pauses before the first character mentioned in the explanation predictors of its pronominalization?*

The key question is whether pronominalization is related to planning. If it is not, then pause length should have no effect on referential form choice. However, if continuations preceded by shorter pauses have more pronominalization, this may indicate that the mechanism by which semantic role predictability in emotion events has an effect on referential form choice is through facilitation of production. This would counter the findings for transfer events (Rosa & Arnold, 2017), and would further support that how we think and talk about these two event types, at least across different experimental paradigms, is significantly different. Additionally, any other early acoustic measure related to the semantic role of the target referent (i.e. the continuation condition), should be investigated as an indicator of facilitated production and predictor of referential form choice.

Methods

Participants & Materials

All participants and items included for analysis in Experiment 1 were used for pause analyses here.

Procedures

All Experiment 1 experimental items followed the same structure (also see *Figure 11* below):

Context + Prompt {Character 1 + Emotion Verb + Character 2} + *because…* + Explanation fact

The context, prompt and *because* were all read directly from the computer screen. However, the learned explanation fact had to be recalled from memory. In addition, it was also emphasized to participants that the fact explanation could be stated in their own words and that they were only trying to communicate the gist of the fact. Thus, participants only needed to syntactically plan

their fact explanations, while the semantic role information of the emotion event along with the coherence relation of the prompt to the explanation were already available or given to them.

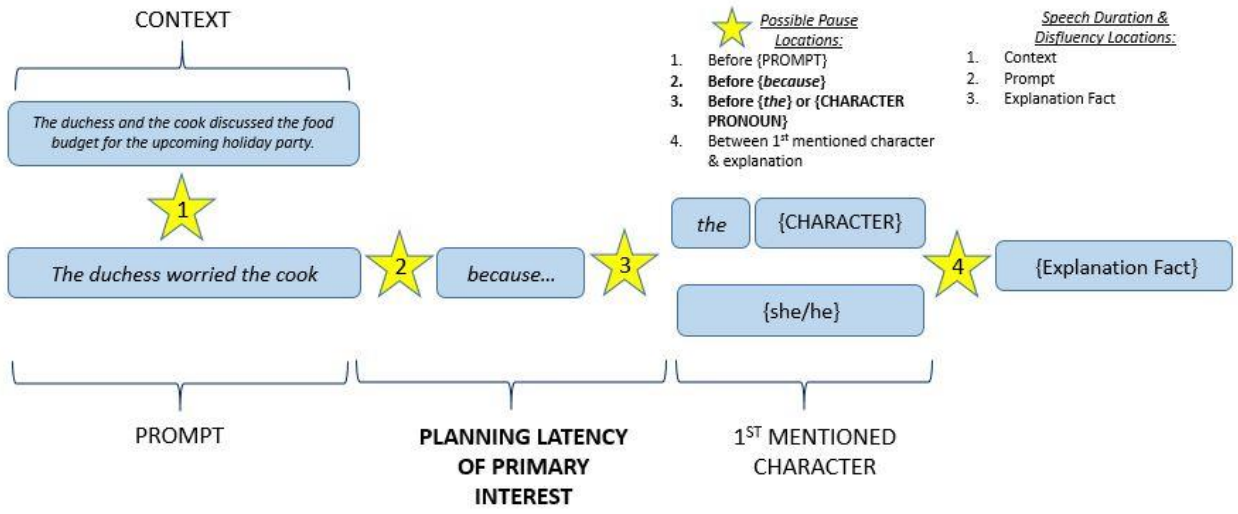


Figure 11: All acoustic measures to be assessed in the critical prompts and continuations.

If latencies reflect planning, and specifically planning difficulty, then greater latency duration (and disfluencies) per trial should be expected for explanations that are either less predictable (i.e. non-cause continuations) or less topical/accessible (i.e. non-subject continuations). This should be primarily true of any pauses, or latencies, between *because* and the explanation fact, but may also apply to a pause between the prompt and *because* and any lengthening of *because (*Figure 10*)*. Therefore the main measure of interest, target latency, was the total latency duration from the offset of the second character in the prompt to the onset of the first character in the fact explanation (the offset of the h/sh/th in *he, she* or the) (Figure 10: summation of 2, "because", & 3). The connective "because" is included as speakers may draw out its pronunciation as they plan the upcoming continuation. Additionally, the duration of pauses prior to the prompt and between the target referent and explanation fact, as well as the duration of the context, prompt and explanation were measured. Disfluencies, defined as stuttering or restarts, were also counted within the context sentence, prompt, and explanation fact.

Audio data coding

All experimental trials included in Experiment 1 analysis were initially included in the latency analysis. Items that exceeded four standard deviations of the overall mean target latency (*M* = 832.79 ms, *SD* = 905.57 ms) or four standard deviations of the overall trial length (*M* = 7497.01 ms, *SD* = 2399.93 ms) were also excluded. Seven items (0.007% of all data) were excluded for exceeding 4 standard deviations from the mean target latency and nine additional items (0.009% of all data) were excluded for exceeding 4 standard deviation of the total trial length. Exclusion of these items did not exclude any additional participants from Experiment 1 for lack of variation in referential form choice.

Experimental trials were coded using Praat (Boersma & Weenink, *praat.org*). Two undergraduate research assistants were trained to annotate and code all included items, marking six time points: (1) the beginning and (2) end of the context sentence, (3) the beginning of the prompt, (4) the offset of the 2nd character in the prompt, (5) the onset of the 1st character in the explanation, and (6) the offset of the fact. Both raters coded the same number of participants' items included for analysis (Participants 1-10) and their measurements for the durations of target latencies were compared for interrater reliability. On average, Rater 1 was 0.73% greater than Rater 2, with a range of 0-9.74% difference. Thus no item differences exceeded 10%. Once it was determined that they were reliably similar, Rater A coded 64.3% of items, while Rater B rated 35.7% of items. Rater A's measurements were used for all double-coded items.

Results

General statistical analyses approach

Logistic regressions were modeled with centered predictors and a maximal random effects structure, using the SAS functions *proc glimmix* (for categorical outcomes) and *proc mixed* (for continuous outcomes). Random intercepts were included for both participant and item if the model converged and was positive definite. Participant by grammatical role

continuation type and Participant by semantic role continuation type slopes were included and retained if the model converged and the matrix was positive definite. Specific model details are included below.

1. *Is target latency related to semantic role or grammatical role?*

Analyses were first performed to determine if grammatical role or semantic role of the continued referent were predictors of target latency. Because the dependent measure was continuous, we modeled target latency with a logistic regression in SAS *proc mixed*. Control predictors from the original Experiment 1 analysis as well as related acoustic measures were first tested in the model, and if significant they were retained. The control predictors tested from Experiment 1 were: participant gender, side correct fact appeared on, list, and if the first character mentioned in the context was also the subject of the prompt. None of these control predictors were found to be significant predictors, as indicated by the dashed lines in *Table 6*. The related acoustic measures included were: context sentence duration, latency before the prompt, prompt duration, target duration, latency before the fact, fact duration, and if there was a disfluency in the context, prompt, or fact explanation. Durations, latencies and disfluencies *prior* to the latency of interest were accounted for as difficulty planning and producing early in the trial could affect downstream planning and representation. Durations, latencies and disfluencies *after* the latency of interest were also accounted for, as these later speech segments may be pre-planned and difficulties might contribute to the latency of interest. Thus, measures of speech production across an utterance are often correlated, as can be seen with several of the control predictors here (*Table 7*). No disfluency or latency measures were significant predictors, and therefore were not retained as indicated by dashed lines in the table below (*Table 6*). All speech duration measures were significant predictors and were thus retained in the final model. Finally, main effects of continuation type on target latency were assessed by adding semantic role, grammatical role and their interaction as predictors in the model.

If cause continuations are easier to talk about, perhaps because they are easier or faster to plan, then we should see shorter target latencies before the cause versus non-cause next-mentioned referent in the explanations, and in fact that is what we find. The target latencies were marginally shorter for cause referents and significantly longer for subject referents (*Figure 12*). These effects were not qualified by an interaction between thematic role and grammatical role. Additionally, the faster all segments of the utterance were spoken, the shorter the target latency was, which may be related to overall speech rates.

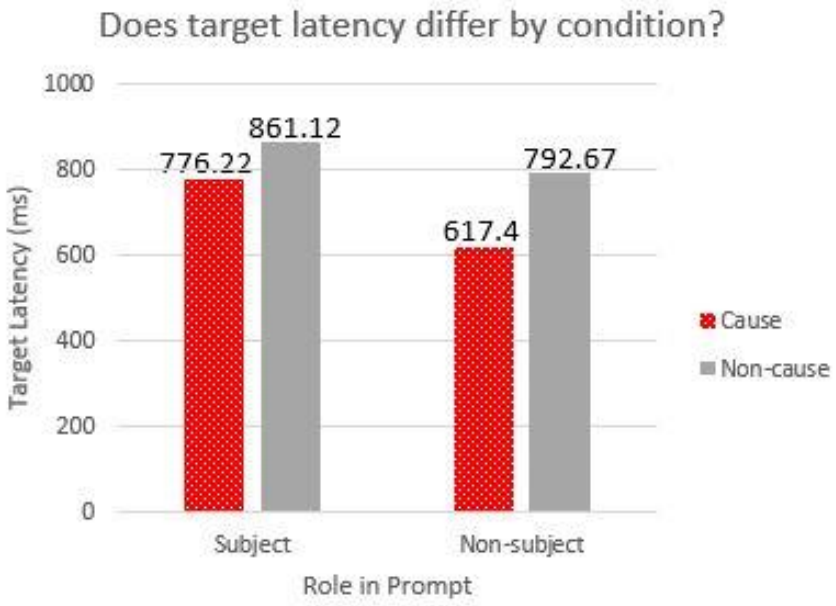


Figure 12. Target latency by grammatical role and semantic role.

DV: Target Latency	Effect	Estimate (St. Error)	*t*	*p*
Critical Predictors	Implicit Cause vs. Non-cause	-0.04 (0.02)	-2.05	0.05*
	Subject vs. Non-subject	0.05 (0.02)	2.74	0.01**
Interaction Terms	Semantic role x Grammatical role	0.002(0.04)	0.04	0.97
Control Predictors	Participant Gender (M/F)	---	---	---
	List (1/2)	---	---	---
	Side Fact Appeared On (R/L)	---	---	---
	1st Context Char. = Subj of Prompt	---	---	---
	Referential form choice	---	---	---
	Context Disfluent	---	---	---
	Prompt Disfluent	---	---	---
	Fact Disfluent	---	---	---
	Context Duration	.22(.07)	3.2	0.003**
	Prompt Duration	.37(.08)	4.89	<0.0001***
	Target Character Duration	.10(.02)	4.7	<0.0001***
	Fact Duration	.14(.04)	3.79	<0.0001***
	Prompt Latency	---	---	---
	Fact Latency	---	---	---
Random Effects	Participant	included		
	Participant by Cause vs. Non-cause	included		
	Participant by Subject vs. Non-subject	---		
	Participant by Context Duration	---		
	Participant by Prompt Duration	---		
	Participant by Target Duration	included		
	Participant by Fact Duration	---		
	Item	included		
	Item by Cause vs. Non-cause	included		
	Item by Subject vs. Non-subject	---		
	Item by Context Duration	---		
	Item by Prompt Duration	---		
	Item by Target Duration	included		
	Item by Fact Duration	included		

Table 6: Predictor variables, control variables, interaction terms and random effects in the final model for target latency to begin speaking the explanation. *Note.* T-values for predictor variables, control variables and interaction terms indicate their significance. Control variables that did not reach significance and were not included in the final model are indicated by dashed lines. Random effects estimated as 0 were also removed from the final model as indicated by dashed lines. Asterisks indicate significance: * marginal /** p<0.05/*** p<0.001.

	Cont. Disfl.	Prompt Disfl.	Fact Disfl.	Cont. Dur.	Prompt Dur.	Exp. Dur.	Prompt Lat.	Exp. Lat.
Context Disfluency	1.00							
Prompt Disfluency	.007	1.00						
Fact Disfluency	.033	.053* (p=.09)	1.00					
Context Duration	**.231** ***	.041	-.010	1.00				
Prompt Duration	.032	**.483** ***	.024	**.341** ***	1.00			
Explanation Duration	-.029	-.011	**.287** ***	.056* (p=.08)	**.116** ** **(p= .0003)**	1.00		
Prompt Latency	.045	-.048	.056* (p=.08)	**.216** ***	**.123** ***	**.010** ** **(p= .002)**	1.00	
Explanation Latency	-.023	.011	**.091** ** **(p= .004)**	-.040	.062* (p=0.05)	**.118** ** **(p= .0002)**	**.162** ***	1.00

Table 7. Correlations between Acoustic Measures

Note. *=p<0.1 **=p< 0.01 ***=p<0.0001

2. *Are other speech or pause measures related to grammatical and semantic role condition?*

Similar analyses were performed to determine if other acoustic measures, both earlier and later in the utterance, were also related to continuation condition (grammatical role and semantic role). Earlier utterance measures included context duration, context disfluency, prompt latency, prompt duration, and prompt disfluency. Later utterance measures included fact latency, fact disfluency and fact duration. Each continuous acoustic measure was modeled with a logistic regression in SAS *proc mixed* (for continuous dependent variables) and with SAS *proc glimmix* for disfluency measures, with semantic role, grammatical role, and their interaction as predictors. Interactions were probed with Bonferroni corrected Least Squares

Means estimates. Because these are all control models, here we only summarize the main effects; all statistical results and specific model details are reported in *Appendix E.*

Earlier Acoustic Measures (prior to Target Referent)

No effect of condition was found to be a significant predictor for the duration of either speech segment prior to the target referent. Thus there were no systematic differences in the length of the context (Appendix E; *Tables 1 & 2*) or the prompt (Appendix E; *Tables 5 & 6*) between trials that continued with causes versus non-causes or subjects versus non-subjects. Additionally, in neither case was the absence of main effects qualified by an interaction of semantic role and grammatical role.

Second, it was determined that disfluency in either speech utterance prior to the target referent did not differ by condition. This was done by testing semantic role and grammatical role as predictors of disfluency of each segment of speech individually (SAS *proc glimmix*). No effect of condition was found to be a significant predictor of disfluency in either segment of speech (Appendix E; Context: *Tables 3 & 4* and Prompt: *Tables 7 & 8*). Additionally, in neither case was the absence of main effects qualified by an interaction of semantic role and grammatical role.

Third, it was determined that prompt latency did not differ by condition. This was done by modeling semantic role and grammatical role as predictors of prompt latency duration (SAS *proc mixed*). No effect of condition was found, and the absence of main effects was not qualified by an interaction of semantic role and grammatical role (Appendix E; *Tables 9 & 10*).

Later Acoustic Measures (after the Target Referent)

Semantic role and grammatical role were also investigated as predictors of explanation fact duration (Appendix E, *Tables 11 & 12*), explanation fact disfluency (Appendix E, *Tables 13 & 14*) and explanation latency (Appendix E, *Tables 15 & 16*). No differences were found for explanation fact duration or disfluency as a result of continuation condition type.

The only significant effect we identified was an effect on explanation latency of semantic role condition qualified by a significant interaction between semantic role and grammatical role

condition was found, such that non-subject non-causes had longer latencies than subject non-causes, and cause non-subjects had shorter latencies than non-cause non-subjects (*Figure 13*), mirroring the pronoun production results of Experiment 1. That is, the one condition where speakers produced fewer pronouns (non-cause/ non-subject) is the same condition where people tended to pause more between the referential expression and the rest of the explanation. This could be related to a tendency to pause more after longer expressions than shorter ones (Watson & Gibson, 2004).

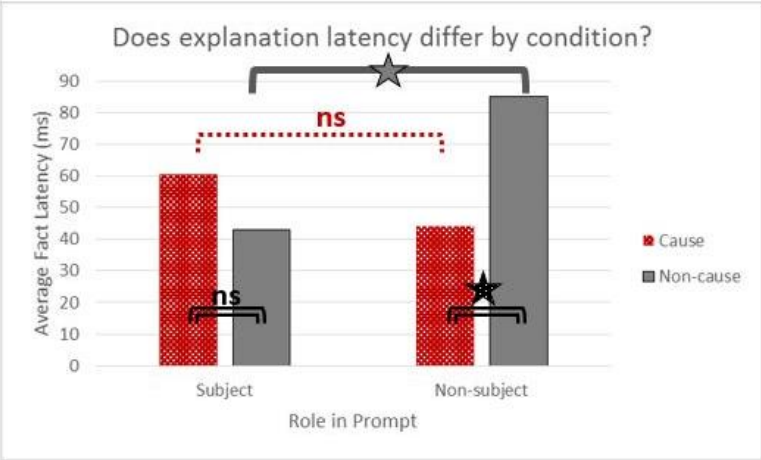


Figure 13. Pause duration between target referent and explanation continuation by condition.

Does production ease and efficiency predict pronoun use?

Of primary interest is whether target latency predicts pronoun use, and if so if this effect eliminates or interacts with the effect of semantic role. Target latency and an interaction between target latency and semantic role were first added as predictors to the final model from Experiment 1 (SAS *proc glimmix*). The maximal random effects model did not converge, and so random effects were removed until the model did converge in the following order: random slope for target latency by item, grammatical role by item, semantic role by item, and random intercept for item. Random slopes for semantic role by participant, grammatical role by participant, and target latency by participant were estimated to be zero, and were removed. Thus, the final model included a random intercept for participant only (*Table 8*).

Critically, there was no significant effect of target latency on pronoun use, and the effect of semantic role remained significant even with the inclusion of target latency as a predictor in the model of pronoun use (*Table 8*). As seen in the previous Experiment 1 analysis, there was a significant effect of grammatical role ($\beta_{\text{grammatical role}}$ = .61(.15), t=4.03, p< 0.0001) and a significant effect of semantic role ($\beta_{\text{semantic role}}$ = .64(.15), t=4.16, p< 0.0001) and these main effects were qualified by a significant interaction ($\beta_{\text{grammatical x semantic role}}$ = -.85(.30), t=-2.81, p= 0.005). Probing the interaction with Bonferroni corrected Least Squares Means estimates found the same pattern of results as seen previously (*Table 9*). More predictable causes were pronominalized more only within non-subjects, while there was no semantic role difference for subjects, and subjects were only pronominalized more for non-causes, while there were no significant differences between grammatical roles for causes. Thus, the pattern of results was similar to the results of Experiment 1 for both semantic role and grammatical role, while no significant effect of target latency was found.

	Effect	Estimate (St. Error)	*t*	*p*
Critical Predictors	Implicit Cause vs. Non-cause	.64 (0.15)	4.16	<0.0001***
	Subject vs. Non-subject	.61 (0.15)	4.03	<0.0001***
	Target Latency	-.15 (0.32)	-0.46	0.65
Interaction Terms	Semantic role x Grammatical role	-.85(.30)	-2.81	0.005**
Control Predictors	Participant Gender (M/F)	---	---	---
	List (1/2)	---	---	---
	Side Fact Appeared On (R/L)	---	---	---
	1st Context Character = Subj of Prompt	---	---	---
Random Effects	Participant	included		
	Participant by Cause vs. Non-cause	---		
	Participant by Subject vs. Non-subject	---		
	Participant by Target Latency	---		
	Item	---		
	Item by Cause vs. Non-cause	---		
	Item by Subject vs. Non-subject	---		
	Item by Target Latency	---		

Table 8: Pronoun use predictor variables including Target Latency, control variables, interaction terms and random effects in the final model of proportion of pronoun use.

Note. T-values for predictor variables and interactions terms indicate their significance. Dashed lines for control variables indicate the variable was not significant in a main effects model and thus was not included here. Random effects are also noted with dashed lines if excluded.

	Estimate (St. Error)	***t***	***p***	**Adj *p***
Non-subjects: Cause vs. Non-cause	-1.07(0.216)	-4.96	<0.0001**	<0.0001**
Subjects: Cause vs. Non-cause	-.221(0.214)	-1.03	0.302	1.00
Non-causes: Subject vs. Non-subject	-1.08(.222)	-4.85	<0.0001***	<0.0001***
Causes: Subject vs. Non-subject	-.229(.207)	-1.10	0.27	1.00

Table 9: Bonferroni corrected estimates for effect of referential predictability of implicit causality for non-subjects vs. subjects and effect of subject-hood for non-causes vs. causes.

Discussion

Speakers are faster to begin talking about explanations that include more predictable implicit causes (at least marginally), but this facilitation of production does not affect referential form choice. Causes do seem to be easier to think and talk about as might be expected of more predictable things, but there is no evidence that this increased ease of production then leads to greater pronoun use. Further, the facilitation of production for predictable causes appears localized to just prior and just after the target referent's production, as neither upstream nor downstream utterances were affected by the continuation condition. While these findings do not support the idea that facilitated production of predictable referents leads to increased pronoun use, the fact that there was facilitated production of the more predictable causes, as well as facilitated production of the rest of the explanation provides evidence that semantic role predictability does reach a great enough threshold early enough to affect at least some aspects of language production. Additionally, there were no differences in the durations or disfluencies of explanations across conditions, which might have suggested confounding experimenter-created systematic differences in the stimuli used. Together with the results of Experiment 2, these

findings support the conclusion that the increased pronoun use for implicit causes in Experiment 1 is due to semantic predictability.

Perhaps counter-intuitively, we also found that speakers took longer to begin uttering continuations with previous subjects than those with non-subjects. It might seem that subjects, as de facto topics, might also be easier to think and talk about. However, subjects were not found to be more predictable in Experiment 2, and previous work with transfer events (Rosa & Arnold, 2017) has found no effect of grammatical role on latencies to begin speaking continuations. This suggests that this finding may be specific to our task.

GENERAL DISCUSSION

This study, using a novel, story re-telling paradigm, offers evidence (1) that people use pronouns significantly more when referring to implicit causes versus non-causes in the emotion events used here, and (2) that implicit causes in these events are more predictable. This finding that more predictable implicit causes are pronominalized more than non-causes differs from previous studies with this event type, but extends previous work with transfer events (Arnold, 2001; Rosa & Arnold, 2017). Importantly, these results are in line with a model of pronoun production in which referential predictability does contribute to referential form choice.

Facilitation of production was also investigated as one possible mechanism through which predictability could affect referential form choice, but no support for this was found. While more predictable implicit causes were faster to plan, this ease of production did not significantly impact the effect of semantic role on pronoun use. This is again in line with previous work with transfer events (Rosa & Arnold, 2017), but leaves open the question of how predictability affects referential form choice for future investigations.

One possibility for determining the mechanism by which referential predictability affects referential form choice is to investigate why the effect is so variable across studies. There are two related avenues for further investigation of this question: (1) why the effect of predictability on pronoun production appears to be so much more stable for transfer events, and (2) why this specific paradigm found an effect with this event type when traditional passage completions have not.

We know that the effect of semantic predictability on referential form is robust for transfer events as compared to emotion events, as the effect with transfer events has been consistently found across passage completion paradigm variations. For example, these

variations have included when the continuation was given or had to be created, as well as with or without supportive contexts, cohesive stories between items and repeated characters between items (Rosa & Arnold, 2017; Zerkle, Rosa & Arnold, 2015). These robust findings indicate that the adaptations of the novel paradigm used here are not necessary for semantic predictability effects to be observed in all event types. One explanation for these disparate findings is that the mechanism by which referential predictability affects referential form choice may actually differ between event types. Thus, future research must be careful not to over generalize any evidence for a specific mechanism in one event type to another by also including investigations of consistency across event types. For example, while facilitated production had not been shown to affect pronoun use with transfer events (Rosa & Arnold, 2017), it was important to also investigate it here as a possible mechanism with emotion events.

A second possible explanation for effect robustness differences between event types is that how we think and talk about these event types may differ in important ways that determine whether, or how much, semantic predictability can affect referential form choice. As previously discussed, transfer events are more concrete, imageable and telic. They also involve a linear and forward movement of events through time (A → B, B→…), whereas the explanations for emotion events might require recalling past events or old knowledge. Emotion events may also require that the speaker model the mental states of at least one, if not both, characters (e.g. why would the butler scare someone/why would the cook be fearful). For any or all of these reasons, emotion events may require more mental resources to represent, leaving fewer resources available to represent the referents involved. Therefore, representationally supportive contexts with strong links between utterances, as with this novel paradigm, may be required in order to see the effects for emotion events. Future studies may manipulate the concreteness, imageability, telicity and/or coherence relation for transfer events in order to systematically investigate if these representational differences matter.

Additionally, explanations for transfer events may generally be more predictable and/or easier to create such that when explanation information is available or created has less of an effect in transfer events. For example, the mechanism by which referential predictability is created in each event type may differ in such a way that predictability information is available at different times or changes across the discourse. Perhaps the novel paradigm used here allows for that information to be available at a time when it can have an effect, unlike in previous studies with implicit causality.

The theoretical debates as to when causality information is available, at least in comprehension, might help illustrate this point. Within the Focusing account, attention is focused on the expected referent as soon as the coherence relation becomes available (Stevenson et al., 1994; Greene & McKoon, 1995; McDonald & MacWhinney, 1995). Within the Integration account, causality information is only available after both utterances have been conceptualized (Garnham, et al., 1996). Therefore, in a traditional passage completion task where participants are reading a prompt (i.e. comprehending) before creating a continuation, causality information might not be available until after the referential form has already been planned, especially under the Integration account (Arnold, 2013). Future research will need to address which, if any, of the paradigm differences used here may have allowed an effect to be seen, and why.

Second, this novel paradigm was designed to address several concerns with the traditional passage completion paradigm. In the traditional paradigm, prompts are typically lacking in substantive contexts. Usually names or various noun phrases are used that give little information from which participants can make representations of the referents. Thus, they may be inventing their own contexts and representations of the referents may vary greatly between producers. Additionally, participants are often required to create their own explanation on the spot. Having sparse context may make creating explanations more difficult and make planning much more incremental, possibly reducing the effects of referential predictability. In the novel paradigm used here the context of each story was highly controlled and supportive of the

explanation. Explanations were also highly controlled and expanded the story without being redundant of earlier information. Additionally, explanations were available to participants prior to having to create its linguistic form which may have encouraged relationships between the events, required fewer mental resources for creating explanations, and mimicked more natural language use. A story-telling context, in which all characters have a visual representation and on-going relationships within a given setting, may have also served to create representational support for these abstract events. Finally, conditions were equal across items, verbs used were highly N1 or N2 biased, all short-stories involved the same six characters who all live and work together in the same setting, and only participants who used some variation in referential form were included. It is beyond the scope of this study, but future research may look at teasing apart which of these aspects is necessary and why.

Additionally, the shorter target latencies before implicit cause referents, in addition to the shorter latencies between the target referent and the rest of the explanation might be the result of increased discourse coherence between the emotion event and the explanations that include more predictable implicit causes. This increased discourse coherence, rather than facilitated production, may allow for a pronoun to be more easily linked back to its anaphor, thus making it more appropriate or easier to use. This could also indicate that the referent is more linked or expected in the explanation that is tied to it. This might indicate that pronouns are used as markers of discourse coherence, and could suggest an audience design or information status mechanism driving referential form choice. Another possible explanation might be that the ease with which an explanation can be recalled, which might be related to how good of an explanation it is, affects the accessibility of the referent. Related to ease of retrieval, the availability of the explanation earlier in this paradigm may also affect accessibility of the referent, and suggests specific reasons to investigate for why this novel paradigm sees an effect of referential predictability. If the explanation is harder to recall, then this may take away from the mental resources needed to represent and plan the referent. In the case of traditional

passage completions where the explanation must be created, this creating process may interfere with resources for representing the referent, concealing the effects of semantic role predictability, especially if this effect is small. Systematically manipulating the coherence between events and their explanations, as well as the difficulty of recall across explanations, may also be additional avenues for future investigation.

Importantly, the effect of predictability on referential form choice found here was not found to be an epiphenomenon of the stimuli used. This was first supported by the findings of Experiment 2 as causes were judged to be more predictable as the next-mentioned character in these specific stimuli. Additionally, there were no significant differences in explanation fact duration or disfluency (measures of production efficiency and proxies for complexity), nor were there differences in the latencies to begin speaking the prompts (a proxy for planning), between items that continued with causes versus non-causes and subjects versus non-subjects. This demonstrates that items were fairly similar in difficulty to produce, at least up until the target referent. This also supports that whatever mechanism semantic predictability affects referential form by has no earlier effects on production in the utterance. For example, while a speaker is producing the prompt they may already be representing and planning the explanation. If that explanation is harder to represent and/or plan, this might make production of the prompt slower or more disfluent. However, no early effects of semantic role predictability or explanation were observed. Importantly, as contexts and prompts were more controlled and consistent between conditions (i.e. subject/cause & non-subject/non-cause items and non-subject/cause & subject/non-cause items had the exact same contexts and prompts that were read), explanation durations did not differ in length between conditions. The explanations could be spoken in the participant's own words, had to be remembered and differed between all conditions. Thus, there may have been greater variability in how they were designed or spoken which may have varied systematically between conditions. However, this does not appear to be the case.

Conclusions and Future Directions

Using a novel story re-telling paradigm, the current study has found, for the first time that this author is aware, that semantic predictability does affect referential form choice for emotion events with implicit causality, suggesting that predictability does contribute to referential form choice, whether directly or indirectly. Extending this effect of referential predictability to an event type in which it has previously not been found is an important contribution to our understanding of referential form choice as it further supports a role for predictability. Future work may investigate the specific aspects of this novel paradigm that contribute to detecting the effect here but not in previous paradigms, or may compare if and how this predictability effect can be manipulated between different event types. These investigations may then inform our understanding of the specific mechanism(s) through which semantic predictability affects referential form choice.

APPENDIX A: CHARACTER DEPICTIONS

the duke

the driver

the butler

the maid

the cook

the duchess

APPENDIX B: EXPERIMENT 1 STIMULI

Cause/Non-cause Verbs (31.1). The implicit cause continuation facts are on the top; Non-implicit-cause continuation facts are on the bottom.

Context Sentence	Prompt	Correct Fact	Incorrect Fact
The duchess and the cook discussed the food budget for the upcoming holiday party.	The duchess worried the cook because…	The duchess always bought the cheapest deal possible.	The duchess had four brothers.
		The cook didn't like to use cheap ingredients.	The cook grew up in Scotland.
The maid and the cook exchanged Christmas presents before work.	The maid pleased the cook because…	The maid was good at picking out presents.	The maid was clumsy.
		The cook loved hand-made presents.	The cook had a goldfish named Vern.
The maid and the duchess were supposed to meet in the kitchen at 9:00 a.m.	The maid annoyed the duchess because…	The maid was always running late.	The maid loved desserts.
		The duchess was always on time.	The duchess was very generous to charities.
The driver and the butler prepared for target practice.	The driver impressed the butler because…	The driver was an excellent marksman.	The driver played the harmonica.
		The butler appreciated good shooting technique.	The butler ran on the track team in high school
The cook and the maid prepared to bake a cake.	The cook delighted the maid because…	The cook was an expert cake decorator.	The cook liked to knit in her spare time.
		The maid wanted to learn how to decorate cakes.	The maid liked to go dancing.
The butler and the driver negotiated a loan and shook hands.	The butler scared the driver because…	The butler had money problems.	The butler liked reading mystery novels.
		The driver hated to lend money.	The driver loved chocolate chip cookies.
The butler and the driver discussed their dreams for the future.	The butler inspired the driver because…	The butler was finishing college.	The butler hated coffee.
		The driver always wanted to become a butler.	The driver was a golf expert.
The duke and the driver stopped to buy cigars on their trip into town.	The duke disgusted the driver because…	The duke loved to smoke cigars.	The duke hated watching television.
		The driver hated cigars.	The driver liked learning to fix things himself.
	The driver amazed the duke because…	The driver was good at reading maps.	The driver had a twin brother.

The driver and the duke took a short cut to town.		The duke was bad with directions.	The duke hated getting dressed up.
The maid and the duchess fought in the kitchen.	The maid offended the duchess because...	The maid did not think before she spoke.	The maid loved traveling.
		The duchess thought staff should never argue.	The duchess loved Italian food.
The maid and the cook gossiped about the Mansion.	The maid amused the cook because...	The maid was a good story-teller.	The maid had blue eyes.
		The cook loved good stories.	The cook always wore a clean, white apron.
The driver and the duke spent the morning fishing, but didn't catch anything.	The driver disappointed the duke because...	The driver hated fishing.	The driver hated talking on the phone.
		The duke wanted to be an avid outdoorsman.	The duke loved to throw big parties.

Non-cause/Cause Verbs (31.2, except "tease" = 31.1). The implicit cause continuation facts are on the top; Non-implicit-cause continuation facts are on the bottom.

Context Sentence	Prompt	Correct Fact	Incorrect Fact
The duke and the driver discussed problems the car was having and possible repairs that needed to be done.	The duke resented the driver because...	The driver was responsible for driving and maintaining the cars.	The driver was very considerate of others.
		The duke hated to spend a lot of money.	The duke was a very confidant person.
The duchess and the cook planned the Christmas party menu.	The duchess admired the cook because...	The cook was very organized.	The cook hated the smell of pickles.
		The duchess was a horrible cook.	The duchess did not know how to drive.
The cook and the duchess often shared tea together in the kitchen.	The cook adored the duchess because...	The duchess was a friendly boss.	The duchess liked going to the theater.
		The cook loved company in the kitchen.	The cook loved purple.
The cook and the duchess looked at family photo albums over tea.	The cook pitied the duchess because...	The duchess didn't have children.	The duchess owned several horses
		The cook had many children and grand-children.	The cook liked to watch soap operas.
The maid and the duchess went over the daily duties of the Mansion.	The duchess envied the maid because...	The maid was just starting out in life.	The maid was absent-minded.
		The duchess was obsessed with looking young.	The duchess volunteered at a local shelter.

The duke and the butler talked about their families.	The butler trusted the duke because…	The duke was a good listener.	The duke loved crossword puzzles.
		The butler did not have a lot of close friends.	The butler was a good Scrabble player.
The maid and the cook put away the dishes on the top shelves.	The cook appreciated the maid because…	The maid was tall.	The maid was fluent in French.
		The cook was very short.	The cook was nosy.
The duchess and the maid talked about how to polish the silver carefully.	The duchess distrusted the maid because…	The maid was a new employee at the Mansion.	The maid was artistic.
		The duchess was a perfectionist.	The duchess always wore her hair up.
The butler and the duke planned to compete against each other in a marathon.	The duke feared the butler because…	The butler had been in the military.	The butler was a bad poker player.
		The duke was out of shape.	The duke loved going to concerts.
The butler and the driver brought in the groceries.	The driver valued the butler because…	The butler was very strong.	The butler liked mystery novels.
		The driver had a bad back.	The driver loved to go camping.
The duke and the butler discussed a pay raise.	The duke liked the butler because…	The butler was a hard-worker.	The butler played guitar.
		The duke appreciated his hard-working staff.	The duke liked to try new things.
The butler and the duke played pool.	The butler teased the duke because…	The duke was not a very good pool player.	The duke was a morning person.
		The butler was a pool champion.	The butler liked hiking.

APPENDIX C: EXPERIMENT 2 (A & B) STIMULI

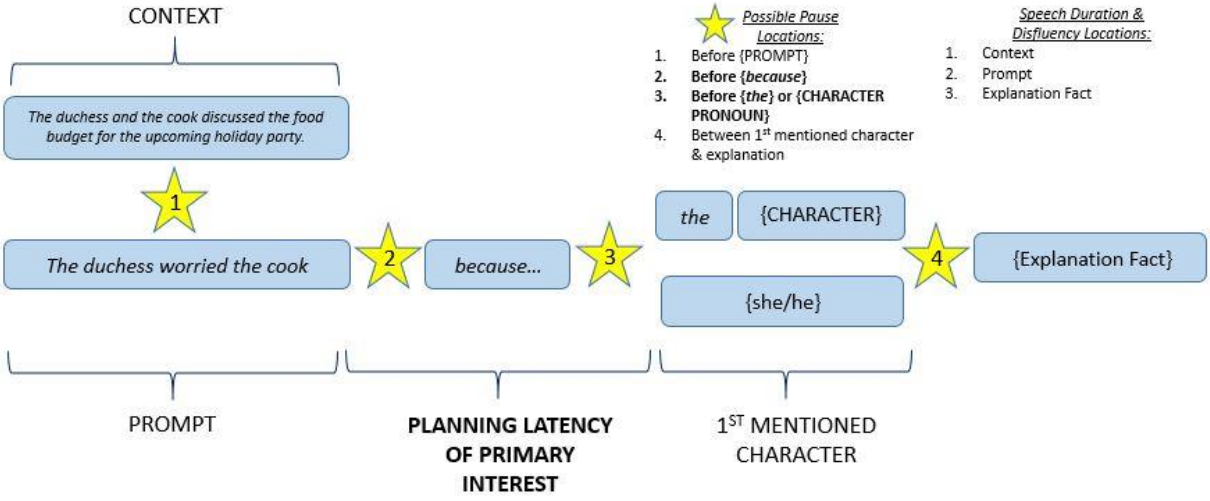
Stimulus-Experiencer Verbs	
Context Sentence	**Prompt**
The butler and the driver discussed their dreams for the future.	The butler inspired the driver because…
The duke and the driver stopped to buy cigars on their trip into town.	The duke disgusted the driver because…
The driver and the duke took a short cut to town.	The driver amazed the duke because…
The maid and the duchess fought in the kitchen.	The maid offended the duchess because…
The maid and the cook gossiped about the Mansion.	The maid amused the cook because…
The driver and the duke spent the morning fishing, but didn't catch anything.	The driver disappointed the duke because…
The duchess and the cook discussed the food budget for the upcoming holiday party.	The duchess worried the cook because…
The maid and the cook exchanged Christmas presents before work.	The maid pleased the cook because…
The maid and the duchess were supposed to meet in the kitchen at 9:00 a.m.	The maid annoyed the duchess because…
The driver and the butler prepared for target practice.	The driver impressed the butler because…
The cook and the maid prepared to bake a cake.	The cook delighted the maid because…
The butler and the driver negotiated a loan and shook hands.	The butler scared the driver because…

Experiencer-Stimulus Verbs	
Context Sentence	**Prompt**
The maid and the cook put away the dishes on the top shelves.	The cook appreciated the maid because…
The duchess and the maid talked about how to polish the silver carefully.	The duchess distrusted the maid because…
The butler and the duke planned to compete against each other in a marathon.	The duke feared the butler because…
The butler and the driver brought in the groceries.	The driver valued the butler because…
The duke and the butler discussed a pay raise.	The duke liked the butler because…
The butler and the duke played pool.	The butler teased the duke because…
The duke and the driver discussed problems the car was having and possible repairs that needed to be done.	The duke resented the driver because…
The duchess and the cook planned the Christmas party menu.	The duchess admired the cook because…

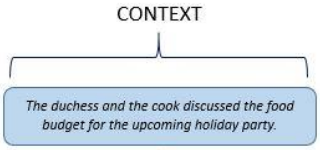
The cook and the duchess often shared tea together in the kitchen.	The cook adored the duchess because…
The cook and the duchess looked at family photo albums over tea.	The cook pitied the duchess because…
The maid and the duchess went over the daily duties of the Mansion.	The duchess envied the maid because…
The duke and the butler talked about their families.	The butler trusted the duke because…

Transfer Verb Fillers	
Context Sentence	**Prompt**
The maid and the cook prepared to serve dinner together.	The cook got a fresh apron from the maid and then…
The duchess and the maid met in the hallway.	The duchess handed the laundry basket to the maid and then…
The duke and the driver shared the morning paper.	The driver gave the sports section to the duke and then…
The driver and the butler washed the cars.	The butler took the bucket from the driver and then…

APPENDIX D: ALL ACOUSTIC MEASURES OF INTEREST



APPENDIX E: RESULTS FOR OTHER ACOUSTIC MEASURES



DV: CONTEXT DURATION

	Effect	Estimate (St. Error)	*t*	*p*
Critical Predictors	Implicit Cause vs. Non-cause	-.005 (0.004)	-0.32	0.75
	Subject vs. Non-subject	-.005(0.006)	-1.33	0.19
Interaction Terms	Semantic role x Grammatical role	-0.02(0.091)	-0.22	0.83
Random Effects	Participant	included		
	Participant by Cause vs. Non-cause	---		
	Participant by Subject vs. Non-subject	---		
	Item	included		
	Item by Cause vs. Non-cause	---		
	Item by Subject vs. Non-subject	---		

Table 1: Context Duration predictor variables, interaction terms and random effects in the final model for context sentence duration.

Note. T-values for predictor variables and interactions terms indicate their significance. Random effects are noted with dashes if not included.

	Estimate (St. Error)	***t***	***p***	***Adj p***
Non-subjects: Cause vs. Non-cause	-0.009(0.46)	-0.20	0.84	1.00
Subjects: Cause vs. Non-cause	0.011(.05)	0.24	0.81	1.00
Non-causes: Subject vs. Non-subject	-0.005(.046)	-0.11	.91	1.00
Causes: Subject vs. Non-subject	0.015(.045)	0.33	0.74	1.00

Table 2: Bonferroni corrected estimates for effect of semantic role and grammatical role continuation condition on context duration.

DV: CONTEXT DISFLUENCY

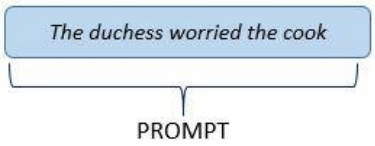
	Effect	**Estimate (St. Error)**	***t***	***p***
Critical Predictors	Implicit Cause vs. Non-cause	-.013 (0.02)	-0.76	0.45
	Subject vs. Non-subject	.003(0.02)	0.17	0.87
Interaction Terms	Semantic role x Grammatical role	-0.02(0.04)	-0.37	0.71
Random Effects	Participant	included		
	Participant by Cause vs. Non-cause	---		
	Participant by Subject vs. Non-subject	---		
	Item	included		
	Item by Cause vs. Non-cause	---		
	Item by Subject vs. Non-subject	---		

Table 3: Context Disfluency predictor variables, interaction terms and random effects in the final model for context sentence disfluency presence.

Note. T-values for predictor variables and interactions terms indicate their significance. Random effects are noted with dashes if not included.

	Estimate (St. Error)	***t***	***p***	***Adj p***
Non-subjects: Cause vs. Non-cause	-0.011(0.03)	-0.41	0.68	1.00
Subjects: Cause vs. Non-cause	0.004(.03)	0.17	0.87	1.00
Non-causes: Subject vs. Non-subject	0.005(.03)	0.17	0.86	1.00
Causes: Subject vs. Non-subject	0.02(.03)	0.77	0.45	1.00

Table 4: Bonferroni corrected estimates for effect of semantic role and grammatical role continuation condition on context disfluency.



DV: PROMPT DURATION

	Effect	Estimate (St. Error)	*t*	*p*
Critical Predictors	Implicit Cause vs. Non-cause	-.004 (0.007)	-0.56	0.58
	Subject vs. Non-subject	.003(0.006)	0.55	0.59
Interaction Terms	Semantic role x Grammatical role	0.013(0.044)	0.30	0.77
Random Effects	Participant	included		
	Participant by Cause vs. Non-cause	included		
	Participant by Subject vs. Non-subject	---		
	Item	included		
	Item by Cause vs. Non-cause	---		
	Item by Subject vs. Non-subject	included		

Table 5: Prompt Duration predictor variables, interaction terms and random effects in the final model for prompt duration.

Note. T-values for predictor variables and interactions terms indicate their significance. Random effects are noted with dashes if not included.

	Estimate (St. Error)	***t***	***p***	***Adj p***
Non-subjects: Cause vs. Non-cause	0.004(0.02)	0.17	0.86	1.00
Subjects: Cause vs. Non-cause	-0.009(.02)	-0.41	0.69	1.00
Non-causes: Subject vs. Non-subject	0.011(.02)	0.46	0.65	1.00
Causes: Subject vs. Non-subject	-0.002(.02)	-0.11	0.92	1.00

Table 6: Bonferroni corrected estimates for effect of semantic role and grammatical role continuation condition on prompt duration.

DV: PROMPT DISFLUENCY

	Effect	Estimate (St. Error)	*t*	*p*
Critical Predictors	Implicit Cause vs. Non-cause	.011 (0.01)	0.77	0.45
	Subject vs. Non-subject	-.007(0.01)	-0.55	0.58
Interaction Terms	Semantic role x Grammatical role	-0.02(0.04)	0.51	0.61
Random Effects	Participant	included		
	Participant by Cause vs. Non-cause	included		
	Participant by Subject vs. Non-subject	---		
	Item	Included		
	Item by Cause vs. Non-cause	---		
	Item by Subject vs. Non-subject	---		

Table 7: Prompt Disfluency predictor variables, interaction terms and random effects in the final model for prompt disfluency presence.

Note. T-values for predictor variables and interactions terms indicate their significance. Random effects are noted with dashes if not included.

	Estimate (St. Error)	***t***	***p***	***Adj p***
Non-subjects: Cause vs. Non-cause	0.0005(0.02)	-0.02	0.98	1.00
Subjects: Cause vs. Non-cause	-0.02(.02)	-0.86	0.39	1.00
Non-causes: Subject vs. Non-subject	0.018(.02)	0.75	0.45	1.00
Causes: Subject vs. Non-subject	-0.001(.02)	-0.07	0.95	1.00

Table 8. Bonferroni corrected estimates for effect of semantic role and grammatical role continuation condition on prompt disfluency.

DV: PROMPT LATENCY

	Effect	Estimate (St. Error)	*t*	*p*
Critical Predictors	Implicit Cause vs. Non-cause	-.030 (0.03)	-0.87	0.4
	Subject vs. Non-subject	.023(0.04)	0.62	0.55
Interaction Terms	Semantic role x Grammatical role	0.052(0.16)	0.33	0.74
Random Effects	Participant	included		
	Participant by Cause vs. Non-cause	---		
	Participant by Subject vs. Non-subject	included		
	Item	included		
	Item by Cause vs. Non-cause	included		
	Item by Subject vs. Non-subject	---		

Table 9: Prompt latency predictor variables, interaction terms and random effects in the final model for prompt latency.

Note. T-values for predictor variables and interactions terms indicate their significance. Random effects are noted with dashes if not included.

	Estimate (St. Error)	***t***	***p***	***Adj p***
Non-subjects: Cause vs. Non-cause	0.06(0.09)	0.65	0.52	1.00
Subjects: Cause vs. Non-cause	0.004(.09)	0.05	0.96	1.00
Non-causes: Subject vs. Non-subject	0.006(.09)	0.06	0.95	1.00
Causes: Subject vs. Non-subject	-0.05(.09)	-0.54	0.59	1.00

Table 10: Bonferroni corrected estimates for effect of semantic role and grammatical role continuation condition on prompt latency.

DV: EXPLANATION FACT DURATION

	Effect	Estimate (St. Error)	*t*	*p*
Critical Predictors	Implicit Cause vs. Non-cause	-.05 (0.03)	-1.65	0.11
	Subject vs. Non-subject	.01(0.03)	0.36	0.72
Interaction Terms	Semantic role x Grammatical role	.12(0.08)	1.59	0.13
Random Effects	Participant	included		
	Participant by Cause vs. Non-cause	included		
	Participant by Subject vs. Non-subject	---		
	Item	Included		
	Item by Cause vs. Non-cause	---		
	Item by Subject vs. Non-subject	included		

Table 11: Explanation Duration predictor variables, interaction terms and random effects in the final model for explanation fact duration.

Note. T-values for predictor variables and interactions terms indicate their significance. Random effects are noted with dashes if not included.

	Estimate (St. Error)	***t***	***p***	***Adj p***
Non-subjects: Cause vs. Non-cause	0.004(0.02)	1.12	0.27	1.00
Subjects: Cause vs. Non-cause	-0.009(.02)	-1.36	0.13	0.75
Non-causes: Subject vs. Non-subject	0.01(.02)	2.30	0.03	.13
Causes: Subject vs. Non-subject	-0.002(.02)	-0.15	0.88	1.00

Table 12: Bonferroni corrected estimates for effect of semantic role and grammatical role continuation condition on explanation duration.

DV: EXPLANATION DISFLUENCY

	Effect	Estimate (St. Error)	*t*	*p*
Critical Predictors	Implicit Cause vs. Non-cause Subject vs. Non-subject	-.015 (0.02) -.01(0.02)	-0.69 -0.63	0.49 0.54
Interaction Terms	Semantic role x Grammatical role	0.05(0.04)	1.29	0.21
Random Effects	Participant Participant by Cause vs. Non-cause Participant by Subject vs. Non-subject Item Item by Cause vs. Non-cause Item by Subject vs. Non-subject	included included --- included --- included		

Table 13: Explanation disfluency predictor variables, interaction terms and random effects in the final model for explanation fact disfluency presence.

Note. T-values for predictor variables and interactions terms indicate their significance. Random effects are noted with dashes if not included.

	Estimate (St. Error)	***t***	***p***	***Adj p***
Non-subjects: Cause vs. Non-cause	0.04(0.03)	1.39	0.17	0.69
Subjects: Cause vs. Non-cause	-0.008(.03)	-0.29	0.77	1.00
Non-causes: Subject vs. Non-subject	0.04(.03)	1.36	0.18	0.72
Causes: Subject vs. Non-subject	-0.008(.03)	-0.28	0.78	1.00

Table 14: Bonferroni corrected estimates for effect of semantic role and grammatical role continuation condition on explanation disfluency.

DV: EXPLANATION FACT LATENCY

	Effect	Estimate (St. Error)	*t*	*p*
Critical Predictors	Implicit Cause vs. Non-cause	-.02 (0.06)	-2.78	0.012
	Subject vs. Non-subject	-.08(0.06)	-1.30	0.209
Interaction Terms	Semantic role x Grammatical role	0.33(0.11)	3.05	0.006
Random Effects	Participant	included		
	Participant by Cause vs. Non-cause	---		
	Participant by Subject vs. Non-subject	---		
	Item	included		
	Item by Cause vs. Non-cause	included		
	Item by Subject vs. Non-subject	---		

Table 15: Explanation disfluency predictor variables, interaction terms and random effects in the final model for explanation fact disfluency presence.

Note. T-values for predictor variables and interactions terms indicate their significance. Random effects are noted with dashes if not included.

	Estimate (St. Error)	***t***	***p***	***Adj p***
Non-subjects: Cause vs. Non-cause	0.33(0.08)	4.09	0.0002	0.0009
Subjects: Cause vs. Non-cause	0.001(.08)	0.02	0.99	1.00
Non-causes: Subject vs. Non-subject	0.26(.08)	3.13	0.003	0.014
Causes: Subject vs. Non-subject	-0.07(.08)	-0.94	0.36	1.00

Table 16: Bonferroni corrected estimates for effect of semantic role and grammatical role continuation condition on explanation latency.

REFERENCES

Altmann, G.T.M. & Kamide,Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition, 73*, 247–264.

Ariel, M. (1990). *Accessing Noun-Phrase Antecedents*. London: Routledge.

Arnold, J. (1998). Reference form and discourse patterns. Unpublished doctoral dissertation, Stanford University.

Arnold, J. E. (2001). The effect of thematic role on pronoun use and frequency of reference continuation. *Discourse Processes, 31*, 137-162.

Arnold, J. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes, 23,* 495-527.

Arnold, J. E. (2010). How speakers refer: The role of accessibility. *Language and Linguistics Compass, 4*(4), 187-203.

Arnold, J.E. (2013). What should a theory of pronoun interpretation look like? Commentary on Kehler & Rohde (2013): A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics, 39*, 1-2, 59-73.

Arnold, J.E. & Nozari, N. (2017). The effects of utterance planning on pronoun comprehension. *Cognition, 160,* 127-144.

Au, T.K. (1986). A verb is worth a thousand words: The causes and consequences of interpersonal events implicit in language. *Journal of Memory and Language, 25*(1), 104-122.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., and Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America. 113*(2), 1001-1024.

Brennan, S.E. (1995). Centering attention in discourse. *Language and Cognitive Processes, 10*, 137-167.

Brennan, S., & Clark, H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology, 22,* 1482-1493.

Brown, R. & Fish, D. (1983). The psychological causality implicit in language. *Cognition, 14(*3), 237–273.

Caramazza, A., Grober, E., Garvey, C., & Yates, J. (1977). Comprehension of anaphoric pronouns. *Journal of Verbal Learning and Verbal Behavior, 12,* 335-359.

Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N. Li (Ed.), *Subject and Topic*, (pp. 25-56). New York: Academic Press Inc.

Chafe, W. (1994). *Discourse, consciousness, and time*. Chicago: Chicago University Press.

Chang, F., Dell, G.S. & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113*(2), 234.

Clark, H.H. (1996). *Using Language*. Cambridge University Press, Cambridge.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral Brain Science, 36*(03), 181–204.

Crinean, M. & Garnham, A. (2006). Implicit causality, implicit consequentiality and semantic roles. *Language and Cognitive Processes, 21*(5), 636-648.

Dell, G.S. (1987). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*(3), 283-321.

Federmeier, K.D. & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41*(4), 469-495.

Fowler, C.A. & Housum, J. (1987). Talkers' signalling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language, 26*, 489-504.

Fukumura, K. & van Gompel, R.P.G. (2010). Choosing anaphoric expression: Do people take into account likelihood of reference? *Journal of Memory and Language, 62*, 52-66.

Gahl, S., Yao, Y., & K. Johnson. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language, 66,* 789-806.

Garnham, A., Traxler, M.J., Oakhill, J. & Gernsbacher, M.A. (1996). The locus of implicit causality effects in comprehension. *Journal of Memory and Language, 4*, 517-543.

Garvey, C. & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry, 4-5,* 459-464.

Givón, T. (1976). Topic, pronoun and grammatical agreement. In C. Li (Ed.), *Subject and topic.* New York: Academic Press.

Givón, T. (1983). Topic continuity in discourse: an introduction. In T. Givón (ed.), *Topic Continuity in Discourse: A Quantitative Cross-Language Study*, 1-41. Amsterdam: John Benjamins.

Gundel, J. (1974). *The role of topic and comment in linguistic theory*. PhD, University of Texas at Austin.

Gundel, J.K. (2010). Reference and accessibility from a Givenness Hierarchy perspective. *International Review of Pragmatics, 2,* 148-168.

Gundel, J.K., Hedberg, N., & Zacharski, R. (1983). Cognitive status and the form of anaphoric expressions in discourse. *Language, 69*, 274-307.

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.

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.

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research, 1626,* 118-135.

Huettig, F., Rommers, J. & Meyer, A.S. (2011). Using the visual world paradigm to study language processing: a review and critical evaluation. *Acta Psychologica, 137*, 151–171.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution.* Oxford University Press.

Jaeger, T. F. (2006). Redundancy and Syntactic Reduction in Spontaneous Speech. PhD thesis, Stanford University, Stanford, CA

James, W. (1890).*The Principles of Psychology*. Henry Holt.

Jurafsky, D., Bell, A., Gregory, M., Raymond, W.D. (2000). Probabilistic relations between words: evidence from reduction in lexical production. In Bybee, J. & Hopper, P. (Eds.), *Frequency and the emergence of linguistic structure.* Amsterdam: John Benjamins.

Kahn, J. M. and Arnold, J.E. (2012). A processing-centered look at the contribution of givenness to durational reduction. *Journal of Memory and Language, 67*(3), 311–325.

Kaiser, E., Li, D.CH., & Holsinger, E. (2011) Exploring the lexical and acoustic consequences of referential predictability. In: Hendrickx, I., Lalitha, D.S., Branco, A., Mitkov, R. (eds) *Anaphora Processing and Applications.* DAARC 2011. Lecture Notes in Computer Science, vol 7099. Springer, Berlin, Heidelberg.

Kaiser, E. & Trueswell, J.C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition, 94*(2), 113–147.

Kehler, A. (2002). *Coherence, Reference, and the Theory of Grammar*. Stanford, CA: CSLI Publications.

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

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

Kehler, A. & Rohde, H. (2014). Pronominal reference and inferred explanations: a Bayesian account. *RefNet Workshop on Psychological and Computational Models of Reference Comprehension and Production*, Edinburgh, UK.

Kutas, M. & Hillyard, S.A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203-205.

Lam, T.Q., & Watson, D.G. (2010). Repetition is easy: Why repeated referents have reduced prominence. *Memory & Cognition, 38,* 1137–1146.

Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. Cambridge: The MIT Press.

Levy, R. and Jaeger, T.F. (2007). Speakers optimize information density through syntactic reduction. In *Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems*.

Piantadosi, S.T., Tily, H.J., Gibson, E. (2009). The communicative lexicon hypothesis. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society,* 2582-2587.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences of the United States of America, 108*(9), 3526–3529.

Pickering, M.J. & Garrod,S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral Brain Science, 27*(2), 169–190.

Pickering, M.J. & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Science, 11*, 105–110.

Prinz, W. (2006). What re-enactment earns us. *Cortex, 42*(4), 515–517.

Reinhart, T. 1981. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica 27*(1), 53–94.

Rosa, E. C., & Arnold, J. E. (2017). Predictability affects production: Thematic roles can affect reference form selection. *Journal of Memory and Language, 94*, 43-60.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928.

Sebanz, N. & Knoblich, G. (2009).Prediction in joint action: what, when, and where. *Topics in Cognitive Science, 1,* 353–367.

Sebanz, N., Knoblich,G., & Prinz,W. (2003). Representing others' actions: just like one's own? *Cognition, 88,* B11–B21.

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

Stevenson, R.J. & Nelson, A.W.R. (1995). The role of parallelism in strategies of pronoun comprehension. *Language and Speech, 38,* 393-418.

Tily, H. & Piantadosi, S.T. (2009). Refer efficiently: Use less informative expressions for more predictable meanings. In *Proceedings of the workshop on the production of referring expressions: Bridging the gap between computational and empirical approaches to reference.*

Van Berkum, J.J.A., Brown, C.M., Zwitserlood, P., Kooijman, V., Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31*, 443-467.

van Rij. J., van Rijn, H., & Hendriks, P. (2012). How WM load influences pronoun interpretation. In N. Rußwinkel, U. Drewitz & H. van Rijn (eds.), *Proceedings of the 11th International Conference on Cognitive Modeling*, Berlin: Universitaetsverlag der TU Berlin.

von Helmholtz, H. (1924). Helmholtz's treatise on physiological optics, Vol. 1, Trans. from the 3rd German ed. (J. P. C. Southall, Ed.).

Watson, D., Arnold., J. E., & Tanenhaus, M. K. (2010). Corrigendum to Tic tac TOE: Effects of predictability and importance on acoustic prominence in language production. *Cognition, 114*, 462-463.

Watson, D. & Gibson, E. (2004) The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, 19, 713-755.

Zipf, G. (1936) The Psychobiology of Language. London: Routledge.

Zipf, G. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. New York: Hafner.