

SEMANTIC ROLE PREDICTABILITY AFFECTS REFERENTIAL FORM

Elise C. Rosa

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in  
partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department  
of Psychology (Cognitive)

Chapel Hill  
2015

Approved by:

Jennifer E. Arnold

Laura Castro-Schilo

Bruno Estigarribia

Kelly S. Giovanello

Peter C. Gordon

©2015  
Elise C. Rosa  
ALL RIGHTS RESERVED

## **ABSTRACT**

Elise C. Rosa: Semantic Role Predictability Affects Referential Form  
(Under the direction of Jennifer E. Arnold)

Referential form variation, such as the choice between *she* and *Sally*, is an important component of meaningful language use. Speakers generally use reduced forms, such as pronouns or zeros, to refer the person who is the topic of the conversation, and who can be assumed to be in the center of attention. Speakers use more specific forms such as names to introduce people for the first time, or to talk about someone who hasn't been mentioned recently (Arnold, 1998; Brennan, 1995; Givon, 1983). The current set of studies examined whether predictability of being referred to also affects referential form. Some (Arnold, 2001) argue that more predictable referents are referred to with reduced forms, but others (Fukumura & van Gompel, 2010; Rohde & Kehler, 2014) argue predictability does not play a role in determining referential form. The current studies manipulated the predictability of pairs of characters in computerized and in-person story-continuation paradigms. Predictability was manipulated using Goal-Source verbs, which make the referent that was in the Goal position more predictable than the Source. In three experiments speakers used more reduced referring expressions when talking about the Goal referent as compared to the Source, in addition to the expected Subject effect. These results suggest that both predictability and topicality information influence referential form.

*This document is dedicated to my parents, Paul Policastro and Patricia Rosa.*

## **ACKNOWLEDGEMENTS**

I would like to acknowledge that my journey through graduate school was not accomplished independently, and I have many people to thank. The first is my advisor, Jennifer Arnold, for providing constant feedback, support, and intellectual prodding, which I am grateful for. Another is Peter Gordon, a member of my committee, who made the important contributions of humor and perspective. I would also like to thank the other members of my dissertation committee, Laura Castro-Schilo, Bruno Estigarribia, and Kelly Giovanello, for their thoughtful comments and encouragement.

I also appreciate the members of the lab who preceded me, Jason Kahn and Alex Christodoulou, whose insight and intelligence contributed to the success of my projects and my approach to many intellectual debates. I am thankful for Cassie Ford and Renske Hoedemaker, the other members of my cohort, who played an important role in my happiness and productivity throughout graduate school. Without Cassie I would not know how to knit, and my life would be far less rich. Without Renske, my officemate of five years, I would have struggled through many a tough day alone.

Finally, but most importantly, I want to thank my family, Paul Policastro, Patricia Rosa, Justine Vlasto, Joey Vlasto, and Leo Vlasto. I know that you love me very much regardless of any achievements, and are as supportive of my Paint art as my dissertation work. However, your support through this process has been critical for me.

## TABLE OF CONTENTS

LIST OF TABLES .....	xi
LIST OF FIGURES .....	xiii
CHAPTER 1: INTRODUCTION .....	1
General introduction .....	1
Predictability at other levels of language production .....	8
Topicality .....	10
Coherence Relations .....	14
Attention-based accessibility mechanisms .....	17
Methods of testing predictability effects.....	20
Current Methods .....	26
Summary and project goals.....	29
CHAPTER 2: TEST OF SEMANTIC ROLE EFFECTS ON PRONOUN/ZERO USE...32	32
General study design .....	32
General analytic approach .....	34
Experiment 1: Online Designated Completion .....	36
Motivation.....	36
Method .....	36
Participants.....	36
Materials and Design .....	37

Procedure .....	37
Analysis.....	38
Response Coding .....	38
Gender Makeup of Characters .....	39
Coherence Relations Coding.....	39
Statistical Modeling .....	40
Results .....	41
Semantic and grammatical role effects on pronoun/zero production .....	41
Discussion .....	44
<b>CHAPTER 3: VARIATIONS ON METHODOLOGY .....</b>	<b>46</b>
Experiment 2: Online Renamed Items Designated Sentence Completion .....	46
Motivation.....	46
Method .....	47
Participants.....	47
Materials and Design .....	47
Procedure .....	48
Analysis.....	48
Response Coding .....	48
Gender makeup of characters.....	49
Statistical Modeling .....	49
Results .....	49

Semantic and grammatical role effects on pronoun/zero production .....	49
Discussion .....	53
Experiment 3: Online Free Sentence Completion .....	54
Motivation.....	54
Method .....	54
Participants.....	54
Materials and Design .....	55
Procedure .....	55
Analysis.....	55
Response Coding .....	55
Gender makeup of characters.....	57
Statistical modeling.....	57
Results .....	58
Semantic and grammatical role effects on pronoun/zero production .....	58
Counts of reference in each condition .....	61
Discussion.....	62
Discussion of Experiment 1-3 .....	63
CHAPTER 4: FURTHER ANALYSIS OF PREDICTABILITY .....	66
Experiment 4: In-person Designated Event Retelling .....	66
Motivation.....	66
Method .....	67

Participants.....	67
Materials and Design .....	67
Procedure .....	68
Analysis.....	70
Response coding .....	71
Audio data coding .....	71
Gender makeup of characters.....	72
Statistical modeling.....	72
Results .....	73
Semantic and grammatical role effects on pronoun/zero production .....	73
Discussion.....	84
Experiment 5: Event Predictability .....	87
Motivation.....	87
Method .....	88
Participants.....	88
Materials and Design .....	88
Procedure .....	90
Analysis.....	90
Results .....	91
Semantic and grammatical role effects on ratings .....	91
Discussion .....	94
Experiment 6: Referent Predictability .....	96

Motivation.....	96
Method .....	96
Participants.....	97
Materials and Design .....	97
Procedure .....	99
Analysis.....	100
Results .....	100
Semantic and grammatical role effects on ratings .....	100
Discussion .....	102
CHAPTER 5: GENERAL DISCUSSION .....	104
Evaluation of Mechanisms .....	110
Conclusions and Future Directions .....	111
APPENDICES	
A. Stimuli from Experiment 1 .....	113
B. Clause coding schema .....	116
C. Stimuli from Experiment 2 .....	118
D. Stimuli from Experiment 3 .....	121
E. Continuation coding schema from Experiment 1 (Rohde).....	124
F. Transcript of background video for Experiments 4, 5, & 6 .....	127
G. Post-experiment questionnaire from Experiment 4 .....	128
H. Stimuli from Experiment 5 .....	129
REFERENCES .....	132

## LIST OF TABLES

### Table

1. Experiment 1 predictor variables, control variables and random effects in the main effects model for referential form choice .....	43
2. Experiment 1 predictor variables, control variables, interaction terms, and random effects in the interaction model for referential form choice .....	44
3. Experiment 2 predictor variables, control variables and random effects in the main effects model for referential form choice .....	52
4. Experiment 2 predictor variables, interaction terms, and random effects in the interaction model for referential form choice .....	53
5. Experiment 3 predictor variables, control variables and random effects in the main effects model for referential form choice .....	60
6. Experiment 3 predictor variables, control variables, interaction terms, and random effects in the interaction model for referential form choice .....	61
7. Experiment 4 predictor, control variables, and random effects in the main effects model for referential form choice .....	75
8. Experiment 4 predictor variables, control variables, interaction terms, and random effects in the interaction model for referential form choice .....	76
9. Experiment 4 predictor, control variables and random effects in the main effects model for connective use .....	78
10. Experiment 4 predictor variables, control variables, interaction terms, and random effects in the interaction model for connective use .....	79
11. Experiment 4 predictor, control variables, and random effects in the main effects model for latency to begin speaking .....	81
12. Experiment 4 predictor variables, control variables, interaction terms, and random effects in the interaction model for latency to begin speaking .....	82

13. Correlations between predictors .....	83
14. Experiment 5 predictor and control variables in the main effects model for Relatedness .....	92
15. Experiment 5 predictor variables, control variables and interaction terms in the interaction model for Relatedness .....	93
16. Experiment 5 predictor and control variables in the main effects model for Predictability .....	94
17. Experiment 5 predictor variables, control variables, and interaction terms in the interaction model for Predictability .....	94
18. Summary of the critical and control predictor effects on referential form from Experiments 1-4 .....	104

## LIST OF FIGURES

### Figure

1. Example of thematic role bias .....	7
2. Sample experimental item .....	33
3. Sample trial from Experiment 1 .....	37
4. Proportion of pro/zero used by semantic and grammatical roles in prior sentence in Experiment 1 .....	42
5. Proportion of pro/zero used by semantic roles and continuation type in prior sentence in Experiment 1 .....	42
6. Sample trial from Experiment 2 .....	48
7. Proportion of pro/zero used by semantic and grammatical role in prior sentence in Experiment 2 .....	50
8. Proportion of pro/zero used by semantic roles and gender makeup of characters in prior sentence in Experiment 2 .....	51
9. Sample trial from Experiment 3 .....	55
10. Proportion of pro/zeros used by semantic and grammatical roles in prior sentence in Experiment 3 .....	59
11. Proportion of pro/zero used by semantic roles and gender makeup of characters in prior sentence in Experiment 3 .....	59
12. Count of total references by grammatical and semantic roles in the prior sentence in Experiment 3 .....	62
13. Sample trial from Experiment 4 .....	68
14. Experimental set-up from Experiment 4.....	69
15. Proportion of pro/zero used by semantic and grammatical roles in prior sentence in Experiment 4 .....	74
16. Latency to begin speaking by grammatical and semantic roles in prior sentence in Experiment 4 .....	80

17. Subject-Goal continuation version of an item from Experiment 5 .....	89
18. non-Subject Goal continuation version of an item from Experiment 5 .....	89
19. Z-score ratings of relatedness by role in prior sentence in Experiment 5.....	92
20. Z-score ratings of predictability by role in prior sentence in Experiment 5 .....	93
21. List A Versions 1 and 2 of an item from Experiment 6 .....	98
22. List B Versions 1 and 2 of an item from Experiment 6 .....	99
23. Count of likelihood judgments for Experiment 6 by semantic role in the prior sentence .....	101
24. Count of likelihood judgments for Experiment 6 by grammatical role in the prior sentence .....	102

## CHAPTER 1: INTRODUCTION

### General Introduction

The study of referring expressions has received a lot of attention from psycholinguists because it is a crucial part of successful language use. In order to express ideas that are understood, speakers must be able to produce references to people or things in the world that their listeners can comprehend. As an example, imagine that you wanted to tell your friend about a real-life mystery show that you watched this week. In particular, you wanted to explain how creepy one of the characters was. You would probably introduce this character the first time using his full name, then go on to explain his bizarre actions. After your first mention of him, you'd probably use a pronoun to refer to him. Later in the conversation you might even call him '*that creep*', knowing that your addressee would know, from the context, to whom you were referring.

The current study will address the question of what factors influence referential form variation, such as the choice between a name and a pronoun. Specifically, it will examine whether the predictability of referring to a certain person influences how they are referred to. A lot of different features of a conversation could make someone more predictable. Your attention might be drawn to someone as they walk by and you might expect that your friend will comment on that person, given their noisy sneakers. Or, you might expect to hear about a certain professor if your friend starts complaining about a class you know he dislikes.

This study will make use of semantic predictability. We will use a certain type of verbs, discussed in more detail later, to influence the predictability of the two referents introduced in a

single sentence. This class of verbs describes transfer events, and includes words like gave, handed, and sold. In a sentence like “*Bob handed the threatening letter to Sue*”, *Bob* takes the thematic role of Source (as he is the person giving the object) and *Sue* takes the thematic role of Goal (as she is the person receiving the object). People have the expectation that *Sue* will be talked about next more than *Bob* (Stevenson et al., 1994; Arnold, 2001), thus *Sue* is more predictable. We will examine who participants choose to talk about, when tasked with providing a completion to a story, and how they refer to that person. The final goal of the study is to examine potential mechanisms of referential form choice.

Referential form varies, and it varies along fairly systematic lines. The critical question for the current set of studies is what mechanism drives this variation in referential form. As a starting point, it is helpful to begin with the current understanding of how referents are represented and pronouns are selected. The prevailing theory of referential choice is that speakers choose referring expressions based on the availability or accessibility of the referent. Referents that are accessible or available have higher activation on a conceptual level (Gundel et al., 1993; Ariel, 1990; Givon, 1983). One critical point is that these theories are based on the assumption that the term referent refers to a non-linguistic representation of a person, rather than the linguistic form used to refer to them (such as ‘*he*’, or ‘*Bob*’). When listening to a story, reading, viewing a movie, having a conversation, or otherwise forming some representation of an event, it is assumed that people build non-linguistic representations of the characters and actions. Each referent, therefore, is proposed to be represented as a distinct entity.

Reduced forms, such as pronouns, are used when the referent is relatively more accessible, and more specific forms are used when the referent is less accessible (Arnold, 2010; Ariel, 1990; Brennan, 1995; Gundel et al., 1993). Speakers also sometimes choose to use zeros to

refer to more accessible referents, such as in the second clause in this example: “*Bob handed the threatening letter to Sue, and then Ø ran out the door*”. One of the most reliable findings is that grammatical Subject-hood makes referents more accessible (Brennan, 1995), which in turn makes speakers more likely to use a pronoun or zero when referring to that referent in a subsequent sentence. For example, following the sentence “*Bob handed the threatening letter to Sue*”, in which *Bob* is the grammatical Subject, “*Then he laughed maniacally*” is a more likely continuation than “*Then Bob laughed maniacally*”. Reduced forms are also often used for items that have been mentioned many times recently (Givon, 1983; Arnold, 2010). Listeners are sensitive to these patterns: they preferentially interpret pronouns as referring to the Subject of the prior sentence and to the referent that was in a parallel position in the prior sentence (Stevenson et al., 1995). However, theories differ on what factors contribute to the referent’s overall accessibility or activation. Another point of debate is whether there is a binary distinction between the most accessible referent and all others or whether all discourse referents are represented with graded levels of accessibility.

One approach to identifying potential mechanisms that determine referential form is to clarify what contributes to accessibility. Some (Kehler & Rohde, 2014; Fukumura & van Gompel 2010) argue that topicality is all that influences accessibility. Topicality is a term meant to capture what the discourse is about, and on the view of some authors (Kehler & Rohde, 2014; Fukumura & van Gompel, 2010) is determined exclusively based on how referents have featured up to the current point in the discourse. An alternative definition of topicality (Givon, 1983) allows for the likelihood of upcoming information, or predictability, to be a contributor to topicality. Arnold (2001) also suggests that predictability may be a feature of topicality. For the purposes of the current discussion the more narrow view of topicality (Kehler & Rohde, 2014;

Fukumura & van Gompel, 2010), will be contrasted with the potential role of predictability. This definition of topicality will be used as it is the view most often associated with this particular question, and it is the view that draws the clearest distinction between topicality and predictability.

The alternative to the possibility that topicality, narrowly defined, exclusively determines accessibility is that both topicality and predictability affect accessibility. Predictability information concerns the future: how referents may feature in the upcoming discourse. The question then becomes a simple one: does topicality alone affect accessibility, or is it a combination of topicality and predictability?

In order to consider predictability in more detail we can briefly return to the example above, in which you were describing a show you'd watched to a friend. Thinking about predictability makes a lot of sense from the perspective of your listener. Your listener doesn't know what you're going to say next, as they didn't watch the show, but they might expect, from having seen other, similar shows, that certain events will take place. For example, they might have a general expectation that the creepy character will be the perpetrator of the crime. That would be a kind of general predictability calculation about the event as a whole. They may also be making predictions about more discrete references, like anticipating who you will refer to next. If you told your friend "*Bob handed the threatening letter to Sue, and then...*" your friend might have the expectation that *Sue* will be more likely to be mentioned next than *Bob*. These more specific predictability calculations about who will be referred to are the kinds we will be focusing on.

Thinking about predictability from the standpoint of the speaker is less intuitive. In general, if you are talking about some event that you're familiar with, or describing a TV show

you just watched, everything you talk about is 100% predictable to you, regardless of how predictable it is to the listener. However, there are several ways in which the predictability of information may influence production: an audience design account, an information status account, and a production facilitation account.

One possibility is an audience design account, in which speakers take into account their audience's perspective, and model the predictability judgments their addressees are likely to be making. This would perhaps lead to more explicit descriptions of unpredictable events, or otherwise signaling to the listener that the event you're describing is not what they were likely expecting (Aylett & Turk, 2004; Lindblom, 1990).

A second way in which predictability could influence production is by facilitating the production process. When formulating an utterance, speakers retrieve the nonlinguistic representations of their ideas, which are then translated into word forms, whose constituent sounds are then retrieved (Levelt, 1989). Speakers are also tasked with selecting or building a frame to arrange their words in a meaningful way. At the conceptual level, if the events you're describing are predictable, such as someone catching a ball after someone else has thrown it, you likely have many representations of such events in your memory. Such similarities between events may aid in your retrieval of the outcome of the event you're trying to describe during the early stages of production. Predictable events might be easier to conceptualize and form linguistic representations of. Unpredictable events might be more difficult to model conceptually and describe. There is evidence that production-based facilitation leads to acoustic reduction (Kahn & Arnold, 2012), and that information about addressee understanding can facilitate production processes (Arnold, Kahn, & Pancani, 2012). If more predictable referents are more

activated and easier to access, this production-based facilitation could result in more reduced form selection.

A third possibility is related to the information status of the event. Many conceptualizations of information status have been proposed, but at a basic level it refers to whether the information being introduced is new to the discourse, or if it is already known or inferable (Clark & Haviland, 1977; Gundel, 1988; Prince, 1981). Information status has been proposed to account for many choices speakers make during production, such as talking about old information before introducing new (Clark & Clark, 1977), and shortening words upon repeated mention (Fowler & Housum, 1987). Under an information status account, rules about the status of the information, given the context, are what drive form selection. For example, there may be a rule that referents under a certain level of accessibility are introduced with names and referents over that level of accessibility can be referred to with pronouns or zeros.

There are many aspects of the discourse and factors external to the discourse that could affect predictability. This set of studies will focus specifically on the predictability associated with different thematic roles. Thematic roles (or semantic roles) refer to the type of role that the referent plays with respect to the verb, in a given sentence. In a sentence like "*Bob handed the threatening note to Sue*", Bob is defined as having the thematic role of Source, as he is the person from whom the action originated. Sue has the thematic role of Goal, as she is the person who the action was designated toward. These semantic roles have known effects on next mention predictabilities: following the example sentence above, people expect the person in the Goal role (*Sue*) to be mentioned next more often than the person in the Source role (*Bob*) (Arnold, 2001). This set of studies will use Goal/Source verbs to motivate predictability expectations. These

verbs describe events of transfer such as give, take, throw, etc. These verbs all take a referent that fills the Goal position and one that fills the Source position.

The predictability of semantic roles is a good test case for examining predictability's role on referential form because it can be separated from the known effect of topicality. In the example sentence given above, “*Bob handed the threatening note to Sue*”; *Sue* would not be considered the topic, as she was not introduced as the grammatical Subject (Kehler & Rohde, 2013). She is, however, predictable based on her semantic role. Therefore, in this case the two accounts diverge.

Goal-Source verbs are ideal for disentangling the effects of grammatical role and thematic role as the class of verbs is further divided into two types. In one type, the Goal is in Subject position (such verbs include *received*, *got*, *took*) and within the other type; the Source is in the Subject position (verbs like *gave*, *handed*, and *sold* fall into this category). This allows for control of the grammatical role while permitting a test of the thematic role.

Given the example above, “*Bob handed the threatening note to Sue*”, the critical comparison is the rate of pronominalization for *Sue*, as the subject of the next sentence, compared to the rate for another character, *Jamie*, in a sentence like “*Larry got the romantic note from Jamie*”. See Figure 1.

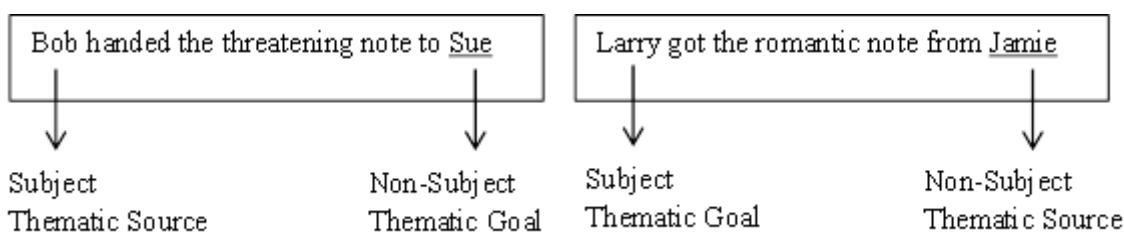


Figure 1. Example of thematic role bias

In the first sentence *Sue* is the non-Subject character, and is the thematic Goal. In the second sentence *Jamie* is the non-Subject character and is the thematic Source. Given the same

grammatical role, is *Sue* pronominalized more, given that she is in the role of Goal, as compared to *Jamie*, who is in the role of Source? The same comparison can be made for *Bob*, who is the grammatical Subject and the thematic Source, and *Larry*, who is the grammatical Subject and the thematic Goal. The effect may be more difficult to detect for the characters in the Subject role, as the strong topicality effects will likely result in high rates of pronominalization for both characters.

The current study will test the first question of whether predictability affects accessibility (and thus referential form) by determining whether the semantic role predictability of Goal/Source verbs influences the form of referring expression the speaker chooses. Some (Arnold, 2001) argue that such verbs bias speakers to refer to the Goal with a pronoun or zero versus a more descriptive referring expression, while others (Kehler et al., 2008; Kehler & Rohde, 2013; Rohde & Kehler 2014; Fukumura & van Gompel, 2010) contend that they do not.

#### Predictability at other levels of language production

Several possible ways in which predictability could affect language production in general were introduced earlier. Briefly, predictability might operate via an audience design account, wherein speakers are sensitive to the predictions being made by their listeners. It might influence production via information status considerations, resulting in unpredictable information being introduced with more description. It could also affect production in a straightforward way by facilitating aspects of production such as planning and word retrieval. There is evidence that if the referent is supported given the context and therefore predictable, the speaker's production process may be facilitated (Gahl et al., 2012; Piantadosi et al., 2009, Tily & Piantadosi, 2009).

A good reason to expect that predictability might affect referential form is that it is known to affect other levels of language production, such as word and vowel duration and final

consonant deletion. Zipf (1936) demonstrated that words that occur more frequently overall, such as function words, tend to be short (for example, *it*, *the*, *and*). Effects of the local context are present as well: Bell et al. (2003) demonstrated that function words that were more predictable in context were produced with shorter vowels than those that were less predictable. Similarly, others (Fowler & Housum, 1987; Piantadosi, 2011) showed that the information contributed by a word, based on its context, is a better determinant of its length than its overall frequency is. Jurafsky et al. (2000) reported that both function and content words were produced with shorter durations and more final consonant deletion when they were more probable. More predictable words, based on overall frequency or local context, are also produced with shorter durations and more final consonant deletion (Gregory et al., 1999). The findings that general and local measures of predictability lead to reduction in acoustic measures also lend support to more general, probabilistic models of language processing (Jurafsky, 1996; Seidenberg & MacDonald, 1999; Hale, 2001).

The Probabilistic Reduction Hypothesis (Bell et al., 1999; Gregory et al., 1999; Jurafsky et al., 1998) formalized the findings of the effects of frequency and predictability on measures of acoustic duration into a hypothesis that states that more predictable words are reduced. This hypothesis allows for many factors to be contributors to predictability, including the context, sentence structure, and discourse and semantic factors. The Uniform Information Density hypothesis (Levy & Jaeger, 2007; Jaeger, 2006) proposed a relationship between the information conveyed by a word and its length. According to this theory, the forms of low-information words are reduced, and the forms of high-information words are lengthened, such that the overall stream of information content is uniform. This hypothesis offers a strategy for dealing with the fact that the total amount of possible information conveyed at any one point is limited, and

distractions, or noise, often compete with the signal during language production. The authors argue that the UID strategy allows speakers to balance two goals: communicating their message while maintaining efficiency.

The UID hypothesis has held at multiple levels of language: overall speech rate (Aylett & Turk, 2004), phonemic production (Son & van Santen, 2005); optional function words production (Jaeger, 2006; Levy & Jaeger, 2007); and contraction production (*you are* vs. *you're*, (Frank & Jaeger, 2008)). At each of these levels, words with high information content are produced in a lengthier manner than words with low information content. When extended to referential expression, another type of language form variation, this hypothesis suggests that less predictable, high-information concepts will be produced with longer forms than highly predictable, low-information concepts. Subjects, and perhaps semantic Goals, therefore, should be referred to with shorter and less informative forms than non-Subjects, and perhaps semantic Sources.

### Topicality

Those who argue that predictability does not influence referential form contend that only topicality information plays a role (Rohde & Kehler, 2014, Fukumura & van Gompel, 2010). Topicality is a term used to describe what a certain discourse is about. The following set of sentences will help illustrate how topicality is established in discourse. Consider the sentences: “*The character in the show, Bob, was really creepy. He was always breaking into people's houses. He used to hide under their beds and stuff. Bob scared Sue out of her own house, and then...*” In this set of sentences *Bob* is the topic, as he is what the sentences are about. There are other people and ideas mentioned, such as *Sue*, but the focus of the passage is on *Bob* and his

behavior. If you were listening to this set of sentences, you would probably have the expectation that *Bob* would be referred to next.

The topicality account (as defined by Kehler & Rohde, 2014; Fukumura & van Gompel, 2010) takes a narrow view: those authors argue that topicality (as represented by grammatical role) is all that influences referential form. Others take a wider view, and define topicality differently. For instance, Givon's (1983) definition of topicality includes information about the past discourse but also about the likelihood of upcoming information (which he calls persistence).

Others (van Rij, van Rij, & Hendriks, 2013) also argue that topicality information, as determined by grammatical role, is the main determinant of referential form, but that working memory plays an important role as well. They found that working memory modulated how well people could attend to the discourse information. For this project we will take a narrow view of topicality for the comparison with a topicality and predictability account, as that is the view most often associated with claims that predictability does not influence referential form.

Proponents of the topicality account (Kehler & Rohde, 2014; Fukumura & van Gompel, 2010) argue that pronoun selection is determined exclusively by this topicality information. For example, given the following, similar, set of sentences "*The character in the show, Bob, was really creepy. He was always breaking into people's houses. He used to hide under their beds and stuff. Bob handed the threatening letter to Sue, and then...*" *Bob* is still clearly the topic of the discourse. However, you may have the intuition that *Sue* will be referred to next (Ferretti et al., 2007; Arnold, 2001, Stevenson et al., 1994), and that reference with a pronoun would be acceptable. The proponents of the topicality account would argue that the reference to *Sue* in this case should not be pronominalized. Under a topicality account, *Bob* can be referred to with a

pronoun, as he was the grammatical Subject. *Sue* was the grammatical non-Subject, and thus a pronoun is not warranted for reference to her.

That is, under a topicality account the semantic features that make reference to *Sue* more likely following the sentence “*Bob handed the threatening letter to Sue, and then...*” as opposed to the earlier one “*Bob scared Sue out of her house, and then...*” should not also warrant pronoun selection. Centering theory (Grosz et al., 1995) provides one framework for categorizing topicality by allowing referents within sentences to be ranked based on whether they have been mentioned previously. Under Centering Theory, pronoun use is based on the ranking of the possible referents. Referents are categorized as either being the backward-looking center, or one of the forward looking centers. The backward-looking center is defined as the referent that is most salient or topical (as determined by grammatical role and previous mention). The forward looking centers are the other referents that were mentioned, and they are ordered based on their grammatical functions. Grammatical subjects are ranked higher than grammatical objects, which are ranked higher than obliques.

Brennan (1995) tested three predictions from Centering Theory by recording participants’ speech while they described a basketball game to someone else. The first prediction was that referents introduced as subjects would be more likely to be continued than those introduced as objects, and would be referents that featured more prominently in the discourse. Second, speakers should only use pronouns for salient entities, such as the backward-looking center. It is worth mentioning that this constraint, among many others of Centering Theory, has been interpreted differently by different researchers. For example, Gordon, Grosz, & Gillion (1993) proposed a stronger interpretation, stating that the backward-looking center should be pronominalized.

The third prediction she tested was that when re-mentioning a referent that was last mentioned as the grammatical non-Subject, speakers should use a full noun phrase (as opposed to a pronoun). She found support for the first claim: speakers introduced characters in high-prominence events as grammatical Subjects, and characters in low-prominence events as grammatical Objects. She found that speakers also used more pronouns to refer to these characters from high-prominence events as opposed to low-prominence events, providing support for the second claim. Finally, when referring to a character that had been introduced as the non-Subject, speakers generally introduced it as a Subject with a full noun phrase before referring to it with a pronoun.

Topicality and grammatical role are tightly linked, particularly within the formal rules of Centering Theory (Brennan, Friedman, & Pollard, 1987). In the final sentence of the first example above, “*Bob scared Sue out of her own house, then...*” *Bob* is also the grammatical Subject, as he is the person performing the action of the sentence. *Sue* is the grammatical non-Subject in that example, as she is the person whom the action is being performed on. However, some have argued that while topicality is usually yoked to grammatical role, it may be a distinct entity. To examine the role of topicality in reference selection, above that of Subject-hood, Rohde & Kehler (2014) manipulated topic-hood of the referent in Subject positions by using a passive construction; (a) *Brittany was amazed by Amanda*, which they argued made *Brittany* highly topical. Additionally, they argued that this construction made *Brittany* more topical than the Subject of a similar sentence; (b) *Amanda amazed Brittany*, allowing for a comparison. Indeed, participants in their studies uttered more pronouns when talking about *Brittany* in sentence (a) than *Amanda* in sentence (b). This does suggest some effect of topicality over and above grammatical position. When referring to the less topical Subject (*Amanda*, in sentence

(b)); participants used pronouns 62% of the time, compared to 86.5% of the time for more topical subjects (*Brittany*, in sentence (a)).

For the purposes of this set of studies, the critical component of topicality, as defined by Kehler & Rohde (2014), and Fukumura & van Gompel (2010), is that it is informed by the grammatical features of the past discourse, and is not informed by next-mention expectations. Any information about the likelihood of a character being referred to that is driven by factors outside of topicality (such as semantic role) is not predicted, by the topicality account, to affect referential expressions.

### Coherence relations

The type of predictability being manipulated within this experiment is reliant upon the discourse. There are other types of predictability that might be less dependent on what had already been said. For example, imagine that you had a friend over, and that you had been talking about several different cats. All of a sudden, your own cat leaped onto your head. Reference to your own cat in the upcoming discourse would become very predictable for your listener, although perhaps nothing in the prior discourse made your cat particularly likely to be mentioned next, among the competitors.

The type of predictability being analyzed in this study is tied to the discourse, and is dependent on the relationships between phrases. Coherence relations is the term used to describe the particular relationships between specific designated sections of discourse (such as sentences) (Kehler, 2002; Stevenson et al., 1994). A coherence relation that is relevant to the current study and class of verbs is a Result relationship. In the sentences “*Bob handed the threatening letter to Sue. Then, she ripped it open*”, the second sentence can be interpreted as happening as a result of the first sentence. In the set of sentences “*Bob handed the threatening letter to Sue. He hoped*

*this would make her pay him the ransom*" an Explanation type of coherence relation relates the second sentence to the first. The second sentence describes the motivating emotion or action that spurred the event.

The coherence relations between phrases modulate semantic role biases for next-event expectations. When the connective focuses on the next outcome, or end state of the relationship, the Goal is preferred. In a sentence pair such as "*Lady Mannerly gave the painting to Sir Barnes. Then...*" speakers are more likely to refer to the Goal of the sentence (*Sir Barnes*), than the Source, and listeners expect the continuation to be about the Goal (Stevenson et al., 1994, Kehler, 2002; Arnold, 2001). If the connective begs an elaboration of the action, or an explanation for it, the Source is preferred. For example, in the sentence "*Lady Mannerly gave the painting to Sir Barnes because...*" *Lady Mannerly* is more often referred to next (Stevenson et al., 1994). Semantic role biases therefore are highly dependent on the coherence relations between phrases.

One issue raised by a consideration of coherence relations is whether speakers and listeners actually have an expectation that the Goal referent will be referred to, or if they are just focusing on the end state of the event. Both these possibilities would lead to the same expectations about continuation. Stevenson et al. (1994; 2000) argued that the Goal bias is an epiphenomenon of the actual focus on the end state of the event. Kehler et al. (2008) found that sentences that described an ongoing event, "*John was handing a book to Bob*", received more Source continuations (referring to *John*) than sentences that described a completed event ("*John handed a book to Bob*"). The Goal and Source of the sentences were identical, so one possible explanation for this pattern of results is that participants were focused on stages of the event itself, rather than one of the referents. Another possibility is that in the version of the sentence

that described an ongoing action, participants' attention was divided between the referents, leading to no strong predictions about either being mentioned next.

Event predictability is the likelihood of possible upcoming actions, while referent predictability is the predictability of a referent itself. We will examine both kinds of predictability within this study, to determine whether participants have expectations about particular upcoming events, and/or whether they have expectations about particular referents. The two are likely not independent, as more predictable events will likely feature more predictable referents. For the purposes of this study we are interested in determining whether semantic role predictability (whether referent or event-based) operates in addition to the known effects of grammatical role/topicality. Our hypotheses are based on that idea that some referents, and the events they are featured in, are more predictable than others. Our stimuli will use perfective verbs to ensure focus is on the Goal/end state of the event.

The choice of referring expressions is expected to reflect the coherence relation between the phrases. In order to determine what effect this has on referring expressions, coherence relations between the sentence prompts and the participants' responses will be examined in the main written story-continuation experiment (Experiment 1). When participants are talking about Goal referents, they may provide more continuations that describe the events that happen next, either temporally or as a direct result of the prompt sentence. When talking about the Source of the action, participants may be more likely to describe the cause of the event or the precipitating actions. When participants provide a role-consistent continuation (e.g., a next-event continuation involving the Goal of the previous sentence), we predict they will use more pronouns or zeros than when they provide a role-inconsistent continuation. This pattern of results would support the importance of coherence relations in facilitating pronoun/zero production.

In the event-retelling experiment the coherence relations between phrases will be controlled, as participants are always describing the next event that happened. This choice was made to utilize the coherence relation that was predicted to support the predictability of goals. If predictability does affect referential form, this design choice should support pronoun/zero use for Goals, and discourage pronoun/zero use for Sources. More natural continuations for Sources might involve talking about the motivation or explanation for an action, rather than describing what happened as a result.

#### Attention-based accessibility mechanisms

If predictability is found to play a role in determining referential form, the second question is by what mechanism. To be consistent with the current understanding of pronoun selection, any mechanism must begin with the assumption that a referent's level of accessibility determines referential form (Chafe, 1976; Prince, 1981; Ariel, 1990; Gundel et al., 1993). The next step in defining potential mechanisms is determining what influences accessibility.

Attention has been hypothesized to be an important determinant of accessibility, and there is some experimental evidence to support this claim (Arnold & Lao, 2015). Within the topicality account attention plays a central role: referents that are being attended to by the speaker and addressee are more accessible and thus more likely to be referred to with pronouns (Chafe, 1994; Levelt, 1989). Within Gundel et al.'s Givenness Hierarchy (1993), which ranks discourse entities based on cognitive status, the referent in the most prominent position is the one that is in the focus of attention.

If predictability also influences referential form, one likely route is through attention. Referents that are likely to be mentioned in upcoming utterances may, in many cases, also be those that attention has been drawn to. One example of such a predictability- based theory is the

Expectancy Hypothesis (Arnold, 1998, 2001, 2008). The Expectancy Hypothesis was developed to provide some framework for quantifying how listeners might make use of topicality information via a psychological mechanism. It was proposed to account for the expectations listeners make during comprehension, but for the current study we will extend it and examine its predictions during production (Arnold, 2010).

The Expectancy Hypothesis states that during comprehension, listeners make predictions about upcoming information, and these predictions drive the accessibility of certain referents in their mental models. Thus, pronouns can be comprehended more quickly when they refer to referents that are more accessible. A corpus analysis demonstrated that speakers are more likely to refer to certain referents based on their semantic and grammatical roles, a pattern of use which may drive the accessibility of certain referents during comprehension (Arnold, 1998; 2001). Under this hypothesis, listeners learn, through language experience, that certain referents (such those that are the grammatical Subjects) are often repeatedly mentioned. This experience eventually leads to the expectation that Subjects, for example, will be repeatedly mentioned, and this in turn contributes to the accessibility of these referents.

The Expectancy Hypothesis allowed for two potential roles of expectancy. Under the strong role, expectancy is the only factor driving accessibility. Under this interpretation there may a straightforward relationship between likelihood of mention and reduced form use, such that highly predictable referents are referred to with pronouns or zeros while less predictable referents are not. The weak role possibility allows expectancy to be just one influence on accessibility, permitting accessibility to be influenced by other factors. This modification is consistent with Givon (1983), wherein many factors, including persistence, or likelihood of upcoming mention, affect accessibility. Furthermore, it is possible that accessibility is modulated

by predictability at different time points, such that a particular referent may have different levels of accessibility at the beginning of a sentence and at the end.

The two mechanisms under consideration that could account for the role of predictability are both based on the assumption that accessibility drives referential form. The first mechanism, which we will call the accessibility mechanism, would allow attention to direct working memory resources to a particular referent, strengthening the accessibility of its representation. Under this mechanism predictability would play a role by modulating attention and directing it (and thus working memory resources) to the referent that was predictable. Information about how topical certain referents are, largely driven by the grammatical context, would also affect attention.

A second possible mechanism, which we will call the facilitation mechanism, would allow topicality to be the sole determinant of accessibility. However, predictability could facilitate other components of the production process, allowing predictable referents to be retrieved and produced more quickly than less predictable referents. Other work has shown that factors that decrease accessibility, such as competition from other characters, also decrease pronoun use (Arnold & Griffin, 2007; Fukumura, van Gompel, & Pickering, 2010; Fukumura, Hyona, Scolfield, 2013). Arnold & Griffin (2007), for example, found that participants were less likely to use reduced forms when there was another animate character present in the picture they were describing, even if they were talking about the most accessible entity. It is reasonable to assume that other factors that lead to decreased accessibility, such as time elapsed since message formulation, would also result in fewer pronouns or zeros.

More specifically, predictability could influence referential form indirectly under this mechanism by reducing the mental resources required to plan and produce the utterance, allowing these resources to be diverted to accessing and representing the discourse context. In

essence, lower working memory resources required for planning the utterance would result in more working memory resources allocated to predictable referents. By contrast, formulating more difficult or less predictable utterances would require more mental resources, draining those available for maintaining representations of the referents (Arnold & Nozari, under review). Another possible role of predictability under this mechanism would be to influence the time course of planning. If predictable information allows for a broader scope of planning, again due to more available resources, then there will be greater discourse connectivity between utterances, potentially leading to more reduced forms.

A third way in which predictability could affect reference production is by an audience design mechanism. This set of studies will not test this mechanism directly, although there is no evidence (Zerkle, Rosa, & Arnold, 2015) that audience design modulates the predictability effect, in this particular task. The current set of studies will focus on distinguishing between the other two possibilities: one in which predictability directly affects accessibility, and one in which it increases pronoun/zero rate by speeding production and planning processes.

#### Methods of testing predictability effects

The current study is designed to resolve a debate in the literature over whether thematic roles affect referring expressions. The theoretical disagreement about the role that predictability plays in determining reference form is accompanied by a contrast in findings. The story-continuation paradigm has been the most frequently used to examine these questions. It provides one method of eliciting information about how people refer and whom they choose to talk about. In standard story continuation studies, participants are presented with a sentence, which they read and then provide a continuation to (Arnold, 2001; Kehler et al., 2008; Fukumura & van Gompel, 2010; Rohde & Kehler, 2014). For example, participants may be presented with a

sentence like “*Lady Mannerly gave a painting of the two of them to Sir Barnes*”. They are then asked to come up with a plausible continuation of this simple story.

One of the outcome measures of interest in such designs is who participants choose to refer to, which differs along two dimensions, grammatical role (Subject/non-Subject) and thematic role (Goal/Source). In the example given, Lady Mannerly is the grammatical Subject and thematic Source, and Sir Barnes is the grammatical non-Subject and thematic Goal. If the sentence were instead worded “*Sir Barnes received a painting of the two of them from Lady Mannerly*” the event described is identical, but Sir Barnes is now the grammatical Subject (but still the thematic Goal), and Lady Mannerly is now the grammatical non-Subject (but still the thematic Source). The critical question in such designs is how participants choose to refer- do they use a name or a pronoun/zero to talk about the character in their continuation?

Arnold (2001) addressed the question of referential form choice using a story-continuation task with 3-sentence prompts. Participants were presented with prompts and then provided continuations out loud. For example, participants read out loud “*There was so much food for Thanksgiving; we didn't even eat half of it. Everyone got to take some food home. Lisa gave the leftover pie to Brendan*”, and then spoke their continuations. She then examined who participants chose to refer to and what expressions they used. She found that participants preferred to talk about Goal versus Source referents in their continuations. More importantly, she found that participants used more pronouns for Goals versus Sources. This effect was primarily driven by the comparison between Goal and Source non-Subjects.

Other methodologies have yielded different results when testing referential form choice in story-continuation tasks. Rohde (2008, Exp. VII) used sentences like “*John handed a book to Mary. ...*” and asked participants to type a continuation. Participants were more likely to provide

Goal continuations than Source continuations. There was not, however, a significant effect of thematic role on pronominalization. Participants did not use more pronouns to refer to Goals in their continuations than Sources. Critically, however, Subject-hood was not manipulated within this study. In order to determine whether a thematic role effect exists, other factors known to affect pronoun use (such as grammatical role) must be carefully controlled and considered. The inconsistent findings to date warrant a closer examination of this question, with the relevant contributing factors controlled. Furthermore, a careful consideration of the particular features of story-continuation paradigms is necessary, given their ubiquitous use in examining these questions.

One benefit of story-continuation paradigms is that they allow for tight control over the experimental materials. The items presented to participants are usually unrelated to one another, to ensure independence across items. Participants build representations of the events they are reading and rapidly construct continuations as they go. This ensures that their continuations are based solely on the linguistic material they're presented with, as opposed to their memories about certain events or how an event actually occurred.

The incremental form of planning that participants must adopt (given that they are coming up with continuations as they go) can be seen both as a benefit and a drawback (see Arnold, 2013, for a critique of these paradigms). It is beneficial in that participants are drawing on the linguistic stimuli as opposed to their past experiences, which cannot be experimentally controlled. It is a drawback in that participants comprehend the sentences for the first time as they're presented with them; leaving relatively few resources to plan their upcoming phrases. They then are tasked with making up an event that could conceivably follow, rather than

describing some known event. This process is likely more mentally taxing than normal language use, and potentially limits the ability of semantic role biases to impact pronoun/zero production.

Although the scope of planning varies during normal language production, people generally have some idea of the upcoming concepts they'd like to introduce. Language production models generally agree that speakers first retrieve non-linguistic representations of concepts, then words, which are fit into sentence frames (Dell 1986; Levelt 1989; Roelofs, 1992). If a set of to-be-described events can be conceptualized as a whole, conceptual relationships will have a better chance of forming between those two events. If, instead, participants produce and utter the descriptions of one event, then come up with a plausible continuation, the two concepts will not have much of a chance of being activated simultaneously. Perhaps instead of viewing the planning constraints as a limitation, the results of these studies should be viewed as more analogous to the mental processes involved in describing an on-going event, such as telling a friend about a football game while you are watching it, or describing a television show while talking to a friend on the phone.

The incremental planning design features of story continuation paradigms might make it more difficult to find the effects of semantic role biases. These biases depend on the conceptual relationships between sentences. Speakers likely make the choice about how to refer to the person in the second sentence of two shortly after they finish producing the first. If the ideas expressed in the two sentences have been activated and conceptualized as a whole, a relationship will exist between the people and actions in the first sentence and the person referred to in the second. This relationship may allow speakers to capitalize on the predictability that the person in the first sentence would be mentioned again, perhaps resulting in a reduced referring expression in the second sentence if the person is very predictable. Alternatively, if the sentences are

conceptualized independently, the predictability of the person in the first sentence being mentioned again will likely not have much bearing on the form used in the second sentence, as no conceptual representation has been developed around the two together (Arnold, 2013). Therefore, the availability of both the prior discourse structure and the upcoming event may be critical for a verb bias effect to be found.

These questions about the time course of the integration of thematic roles are debated in the comprehension literature, within the Integration and Focusing accounts of Implicit Causality information. In sentences like “*Susan praised Diane because she...*” some (Stevenson et al, 1994; Greene & McKoon, 1995; McDonald & MacWhinney, 1995) argue for the focusing account: attention is focused on the referent that is consistent with the coherence relation, as soon as that relation is available. Others (Garnham et al. 1996) argue that the causality relationship is only used later, when integrating the clauses. If thematic roles do indeed affect referential accessibility, one question is what the time course of this effect is. If these effects depend on the coherence relations between phrases, then the accessibility of the referents can't be affected until enough is known about the following clause to establish the coherence relation. If, however, the relationships between clauses are predicted ahead of time, the thematic role information could affect accessibility earlier. Thematic role information therefore is incorporated differently into the discourse model than grammatical role information, which can have a more immediate effect on accessibility. This has implications for production as the grammatical role information can more quickly and easily affect accessibility, driving pronoun/zero use. Thematic role information is incorporated later, and thus its effects may not occur in time to drive form selection.

A final component of the story continuation paradigm that should be examined in terms of its implications for ecological validity is the use of written language. All the studies discussed

above (Kehler et al., 2008; Fukumura & van Gompel, 2010; Rohde & Kehler, 2014; Rohde, 2008) with the exception of (Arnold, 2001) used written or typed responses. It has long been noted that spoken language is targeted at a particular audience, whereas written language is produced without a particular audience in mind (Vygotsky, 1962; Sartre, 1964). Written language has been found to be more complex (Halliday, 1979; Harrell, 1957) and time-intensive (Horowitz & Newman, 1964) than spoken language, which is rated as more interesting and contains a simpler vocabulary (Gibson, 1966; Nida, 1967).

There is also more of a need to be unambiguous when writing, because writers can't rely on cues from their readers to tailor how much information they provide. This need to be more unambiguous may encourage less pronoun use in written versus verbal responses. It is also not possible to make use of prosody, which helps indicate the focus of a sentence, while writing. Additionally, the time course of planning differs between writing and speaking, allowing for more complex constructions to be planned and used in writing versus speaking. Many of the same operations likely underlie the production of written and spoken story continuations. However, allowing participants to provide verbal continuations is more naturalistic and such results may differ in meaningful ways from those obtained via written participation.

With the exception of (Arnold, 2001), studies of semantic role predictability effects on reference form have concluded that the factors that influence likelihood of reference and those that influence referential form are different (Kehler et al., 2008; Kehler & Rohde, 2013; Rohde & Kehler, 2014). That is, while speakers may prefer to speak about the Goal referent, this bias does not result in more reduced forms being used to refer to that person.

Rohde & Kehler (2014) did address several of the limitations of previous work on semantic biases, including the use of nonambiguous context (by using two different-gendered

characters), and de-conflating the roles of topic and Subject-hood. They argued that the previous use of nonambiguous gender contexts was a limitation, as speakers could use a pronoun for the grammatical Subject regardless of predictability manipulations, as the pronoun would be interpretable. Use of ambiguous contexts has a drawback as well, though, as participants could just perform at floor (use no pronouns) for the non-Subject character (again, this would wash out any predictability manipulations). Given this possibility, both gender combinations (same-gender and different-gender) were used in the current experiment.

Despite addressing many previous limitations, some shortcomings remained in Rohde & Kehler (2014). While they were right to assert that the stakes were low in prior studies, given the unambiguous nature of pronoun use, they used a standard story continuation task in which there was no obvious addressee. With no addressee comes relatively few communicative demands, or at least different communicative demands than exist with a live addressee. There is also lower motivation for correct task completion.

### Current methods

For the reasons detailed above, story continuation studies to date may not have been designed optimally to find the effects of semantic role biases. These problems were addressed in several ways. First, we created a modified version of the story-continuation paradigm. This modified version was created to address the critical limitations of previous designs while retaining the valuable controls.

The first modification made to distinguish the current set of studies from prior designs was the adoption of the use of Goal/Source verbs. Most of the prior studies (Fukumura & van Gompel, 2010; Rohde & Kehler, 2014; Kehler et al. 2008 Exp. 3; Kehler & Rohde, 2013) used a class of verbs called Implicit Causality, which differs in important ways from Goal-Source verbs,

used in Arnold (2001). More importantly, these verbs differ in ways that may make effects of verb bias more difficult to detect. Like Goal-Source verbs, Implicit Causality verbs take two referents. One referent takes the Stimulus role and one takes the Experiencer role. In a sentence like *John admired Amy*, *Amy* is the Stimulus and *John* is the Experiencer. The referent who is the Stimulus, *Amy*, is the one responsible for the emotion evoked in the other referent, *John*, the Experiencer. One way in which these verbs differ from Goal-Source is that they describe events that are psychological states (such as *admired* and *blamed*), which are less imageable than discrete actions (*handed*, *gave*). It is possible that the more abstract nature of these verbs and the fact that they describe psychological states as opposed to actions may make them harder to conceptualize and incorporate into mental models, as compared to Goal-Source events. Given that semantic role biases depend on coherence relations between phrases, any difficulty incorporating these events into mental models would necessarily diminish these effects.

Implicit Causality verbs and Goal Source verbs also differ in their telicity. Telicity is a linguistic property of verbs which make a distinction between verbs that describe events which can be conceptualized as having an endpoint (telic verbs) and those that describe events that are conceptualized as lacking endpoints (atelic verbs). Consider the event described by *handed*. This event is conceptualized as having a beginning, the object moving from the Source referent, and an ending, the object being delivered to the Goal referent. The event described by the word *admired*, on the other hand, lacks these distinct beginning and end points. The admiration may have begun at any point in the distant or recent past, and has no distinct endpoint in the future.

Given that telic verbs, those with distinct beginning and endpoints, could potentially be easier to conceptualize, this may allow them to be incorporated into mental models more easily than atelic verbs. Modeling the distinct action in a discourse model might help focus attention on

one of the referents, by drawing attention to either the beginning or end of the event.

Incorporating an atelic event into a discourse model, by comparison, does not include modeling a concrete beginning of the event, associated with one referent, and a distinct endpoint, associated with the other. This could result in attention being more evenly divided across the event participants.

The current study will also adopt different exclusion criteria than were used previously. For the current experiments, only participants who used some variation in their referring expressions (at least two pronouns or zeros and at least two names) could be included. Former studies (Fukumura & van Gompel, 2010; Kehler & Rohde, 2013) did not use this exclusion criterion, leaving open the possibility that some of the participants included were not using any referential variation. If the semantic predictability effect is fairly small, including participants who used no variation would make it even harder to find, and these ceiling or floor performances might mask the thematic role effect. This exclusion criterion is similar to the convention, often used in other studies, of excluding participants who are performing at floor or ceiling on a certain task (Filmer, Mattingly, Dux, 2015; Buschkuehl et al., 2014). In this set of studies the outcome measure is pronoun/zero use, and producing all pronouns/zeros puts participants at ceiling.

Another improvement of the current design was to control for grammatical role. In prior studies grammatical role has not always been controlled (Kehler et al., 2008) or manipulated. Given that Subject-hood is a primary contributor to referential form, this oversight would likely dwarf any other effects that exist in the data.

The current studies will also control and examine for the effects of gender ambiguity on referential form. Half of the experimental items will feature characters of different genders, and half of the items will feature characters of the same gender. We expect that the overall rate of

pronoun/zero use will be lower for same-gender character items. This could occur for a couple of potential reasons. First, perhaps participants are sensitive to the ambiguity of pronouns, and thus will use fewer pronouns when the characters are of the same gender via an audience design constraint. A more production-based explanation is that characters of the same gender compete for activation (Arnold & Griffin, 2007; Fukumura, van Gompel, & Pickering, 2010), as they share some similar semantic features. Perhaps this competition is what diminishes their overall accessibility, leading to lower rates of pronominalization. In either case, it is important to control for the known effects of gender on referential form.

Second, we added the use of an in-person paradigm to answer these same questions in a more naturalistic and interactive setting. The in-person paradigm makes use of the same items as the more traditional sentence-continuation experiments, but incorporates addressee demands and motivations. Two critical components of this paradigm are the focus on the interactive quality of the task and participants' engagement. These are particularly important given the nature of the question: predictability of upcoming information only matters if you are paying attention to the information and using it for a purpose. This particular experiment will allow the participant and confederate to collaborate on a task, with the end goal of solving a mystery.

### Summary and project goals

In summary, the first question this set of studies will address is whether predictability, manipulated in this study by the thematic roles of Goal-Source verbs, affects referential form. Evidence for the role of predictability would come in the form of differences in pronoun/zero use between Goals and Sources. If topicality is the sole determinant of referential form then the rates of pronouns and zeros used for Goals versus Sources should not differ. If predictability increases pronoun/zero use, more pronouns and zeros should be used to refer to Goals versus Sources. If

predictability is found to play a role in determining pronoun/zero use, the second question is by what mechanism. We will focus on the distinction between the referent accessibility and facilitation mechanisms.

The referent accessibility mechanism allows for predictability information to affect accessibility directly, by directing attention and working memory resources to the predictable referents. This study will not be able to provide information that would uniquely point to this mechanism as the correct one, especially since even if it were the true mechanism, predictability would likely also have effects on planning and production. The focus instead is on whether specific information that would uniquely select a facilitation mechanism is found.

The facilitation mechanism under consideration allows for topicality to be the sole determinant of accessibility. Predictability information, however, could facilitate production and planning, speeding retrieval of the utterance. Evidence to support this mechanism would come in the form of planning measures that predicted referential form. Specifically, this mechanism would be supported if we find shorter latencies to begin speaking when pronouns/zeros are used. There might also be a relationship between referential form and other measures of planning difficulty. If fewer pronouns and zeros were used when speakers were disfluent or had trouble retrieving the verb, that finding would be interpreted as support for this mechanism.

The studies will be discussed as follows. First, the improved story-continuation paradigm will be used to determine whether semantic roles affect pronoun/zero production (Exp. 1). Second, slight changes will be made to the design to determine whether an overarching discourse context is necessary to find the effect (Exp. 2), and to confirm that Goals are considered more predictable than Sources, as measured by speakers' preferences to refer to them (Exp. 3). The in-person design (Exp. 4), which provided important time-course information to inform the two

mechanisms being considered, will then be discussed. Finally, more information about event-based and referent predictability, obtained from two ratings studies (Exp. 5 and Exp. 6), will be considered.

## **CHAPTER 2: TEST OF SEMANTIC ROLE EFFECTS ON PRONOUN/ZERO USE**

### General study design

The stimuli for this study were designed with verbs that describe transfer events, called Goal-Source verbs. The objective of this study was to test (1) whether Goal referents are more predictable than Source referents, and (2) whether this predictability leads participants to refer to Goal referents with more reduced referring expressions (such as pronouns or zeros) than Source referents. We used a story-continuation paradigm to test for these effects in three experiments. The sentences used were designed to be usable with an event-retelling study in which the events were depicted as well (Experiment 4). Two rating studies were also completed to answer additional questions about the relatedness and predictability of the materials. Detailed information about each task as well as any modifications from the general experiment design can be found in each experiment's description.

The storyline for the Goal-Source verb studies was developed by the experimenter and featured characters in a Clue-like murder mystery. In this murder mystery there were three main male characters: Sir Barnes, the chauffeur, and the butler, and three female: Lady Mannerly, the chef, and the maid. The characters' behaviors and actions were consistent with their real-world roles.

The storyline was divided into pairs of sentences, which described actions that took place involving two of the characters (in the critical stimuli items) or one to three characters (in the filler items). Two versions of the first sentence in each pair were created. These two versions were identical in content but differed in structure, allowing for the characters that filled the Goal

and Source roles to be consistent while changing whether they were the Subject or the non-Subject of the sentence. For example, “*The maid learned cake decorating from the chef*” versus “*The chef taught cake decorating to the maid*”.

The critical stimuli items were evenly divided between Goal and Source continuations and within those divisions were evenly divided between sentences that described characters of the same gender and characters of different genders. Therefore, there were 12 critical items in which the Goal of the first sentence served as the continuation referent and 12 critical items in which the Source was meant to be continued (in the designated completion versions of the experiment). Within each of those twelve, six of the items had characters of different genders interacting and six had characters of the same gender.

The first sentence in each pair was used as a prompt. The second sentence was used to create pictures shown in later experiments and was also presented to the participants in one of the rating studies. Two sample items, in each condition, are given in Figure 2.

**Sample Item 1:**

Goal continuation, non-Subject position:  
*Lady Mannerly gave a backrub to Sir Barnes.*

Goal continuation, Subject position:  
*Sir Barnes got a backrub from Lady Mannerly.*

**Sample item 2:**

Source continuation, non-Subject position:  
*The maid took a cookbook from the chef.*

Source continuation, Subject position:  
*The chef handed a cookbook to the maid.*

*Figure 2.* Sample experimental item

In the sentence completion paradigms each participants read the 24 critical sentences and 29 filler sentences. In the event-retelling paradigm they heard the sentences and saw them depicted. In the rating studies they read the sentences and saw them depicted. The stimuli were arranged following a Latin Square design, and each participant viewed each target item only once. The use of a Latin Square design allowed each participant to be exposed to each item in only one condition, but see all conditions across different items. This design also helped discourage notice of any pattern of reference across the items. The filler items helped develop the storyline and added variety to the kinds of sentence forms encountered.

The presumed predictability of the intended referent in the second sentence varied along two spectrums. Referents that were the Goal of the first sentence were more predictable than those that were the Source, and those that were the Subject were more predictable than those that were the non-Subject. Referents that were both Goals and Subjects of the first sentence were most predictable, and those that were non-Subjects and Sources were the least predictable. Therefore, the four conditions of the experiment were (1) Goal, Subject reference; (2) Goal, non-Subject reference; (3) Source, Subject reference; and (4) Source, non-Subject reference.

The same participant exclusion criteria applied for all the experiments. In order to be included in the analyses participants needed to use at least two pronouns/zeros and at least two proper names. This qualification ensured that participants were using some variation in their referring expressions and were not performing at floor or ceiling.

#### General analytic approach

Generalized linear mixed-effects models were used to account for the dependencies in the repeated measures and to appropriate model the dichotomous outcome. SAS proc glimmix was used for analyses of dichotomous outcomes, and SAS proc mixed was used for analyses of

continuous outcomes. The same analytic approach was used for all the experiments. Any adjustments to this approach will be discussed in detail in the analysis section of each experiment's description. First, a control model was built with any potential variables that might affect the outcome variable, other than the critical predictors of semantic and grammatical role, gender, and order. Control models were run with a random intercept to account for individual differences among participants. The control models did not include a random effect for items as the models did not converge with this effect. Any control variables that had a t-value of greater than 1.5 were retained for the main effect models. Main effects models and interaction models were then constructed.

The main effect models were built with the relevant control variables (those with  $t$ -values  $> 1.5$  in the control model) and the critical predictors, Subject/non-Subject, Goal/Source, gender, and order. Order was included as a fixed effect. Order was the same as item in this experiment, as the items were presented in the same order to all participants, to preserve the story-nature. Goal/Source, gender, and Subject/non-Subject were centered by coding them as 0/1 and grand-mean centering. Participant was included as a random intercept in the main effect models. Random slopes for participant by Subject/non-Subject and participant by Goal/Source were also included if the models converged and were positive definite. Such details will be given for each analysis.

The main effects of semantic and grammatical role were the focus of the analyses. In order to check whether these effects were qualified by interactions, the final step was to build interaction models for each analysis. The interaction models included only control variables that were significant in the main effects models, the critical predictors that had been tested in the main effects models, and the random effects terms from the main effects models. The interaction

terms Subject\*Goal, Goal\*Gender, Subject\*Gender, and Subject\*Goal\*Gender were tested in the interaction models. These interactions captured the relationships between grammatical and thematic roles, gender of characters and thematic role, gender of characters and grammatical role, and the relationship between gender of characters, thematic role, and grammatical roles. Given the relatively small sample sizes and potentially limited power, any interaction with a *p*-value of <.10 will be considered significant and examined further. In the event of a significant interaction, contrasts were added to the interaction model to test for the specific effects of the interaction. Any other interaction terms in addition to the four described above, specific to a particular model, will be explained in the analysis section of that experiment.

## EXPERIMENT 1: ONLINE DESIGNATED SENTENCE COMPLETION

### *Motivation*

Experiment 1 was conducted using the story-continuation methodology to determine whether participants used more pronouns/zeros when referring to Goals of the prior sentence than Sources. It also examined whether participants used more pronouns/zeros to refer to Subjects of the prior sentence as opposed to non-Subjects, as would be expected.

### Method

#### *Participants*

46 undergraduates completed the task, 10 for course credit and 36 for a monetary reward. In order to be included participants needed to be native English speakers, have normal or correct-to-normal vision, and couldn't have participated in a similar experiment in the lab. Participants also needed to use at least two pronouns or zeros and at least two names in their responses to be included. 20 participants were included in the analysis. Of the 26 who were not included, 8 were excluded for using fewer than 2 proper names, 17 were excluded for using fewer than 2 pronouns

or zeros, and 1 was excluded because we had collected enough data for even numbers on each list.

### *Materials and Design*

The experimental trials in this experiment consisted of the first in each pair of sentences from the storyline. There were 24 critical items and 29 fillers. Participants were recruited through SONA or flyers on campus. The sentences were presented to participants with a computerized survey through Qualtrics. Participants were instructed to provide a plausible continuation about the character that was underlined in the first sentence. Participants were presented with one of the two lists created, allowing them to see each item in one of the two conditions, but both conditions across items. Participants were therefore providing completions about the same characters later depicted in the second picture of each pair of the event-retelling paradigm, with the exception of two of the filler sentences. These filler sentences were modified slightly for this experiment, as they depict (in the second sentence) characters that were not mentioned in the first. For those items in this experiment, one of the characters in the first sentence was indicated as the desired continuation. The stimuli for Experiment 1 are in Appendix A. An example item from Experiment 1 is shown in Figure 3:

Lady Mannerly gave a painting of the two of them to Sir Barnes.

*Figure 3. Sample trial from Experiment 1*

### *Procedure*

Participants were presented with the first sentence of each pair, as described above. Participants were asked to provide a plausible continuation about the character that was underlined. Participants provided typed responses via computer.

## *Analysis*

### *Response coding*

The subjects of the first clauses of participants' responses were coded. The coding schema for determining clauses is given as Appendix B. Briefly, a clause was considered to be a main or subordinate clause and its arguments. The subjects of the first clauses were coded for whom the participant had written about and how they referred to that person. That is, responses were coded for choice of referring expression (pronoun/zero or proper name) and role of referent in the prior sentence (Subject or non-Subject and Goal or Source). Items were excluded if participants referred to more than one person at once (e.g., *Then they put the groceries away*), they did not refer to the character that was underlined, they referred to some inanimate object as the subject, or they referred to someone's possession or body part as the subject (e.g., *Sir Barnes' back was sore*). Participants who did not use at least two pronouns or zeros or at least two names were excluded from the analysis altogether.

The author coded the data and then one of two undergraduate research assistants re-coded the data, blind to the original coding. Of the 480 items there were nine in which the original and re-coding did not match, or 2.10% of the data. In six of those cases the two coders disagreed on who the intended referent was, or one thought it was unclear. All those cases were eventually excluded because the intended referent could not be determined. The other three cases were ones in which the research assistant miscoded or forgot to code the item. For these items the two coders conferred and the original coding was agreed upon.

Fifty-five items were excluded from the final analysis, leaving 425 items. Seven items were excluded for being about non-human referents, one was excluded for referring to multiple characters, one was excluded for being nonsensical, five were excluded for being possessive

Subjects, five were excluded because it was unclear who was being referred to, and 36 were excluded as the wrong character was referred to. Counts of items in each condition were fairly equal. There were 223 Goal items and 202 Source items, 223 Subject items and 202 non-Subject items.

Responses were also coded for use of connectives (*if, then, and, so*, etc), and for whether the non-designated character was referred to in the same clause as the designated referent.

#### *Gender makeup of characters*

The chef, originally intended to be female for the purposes of balancing the gender makeup of the characters, was interpreted as male by most of the participants. The coding of gender for the items in which the chef appeared was changed to reflect this. Another ambiguously gendered character, the sales clerk, was intended to be male but most participants interpreted him as a female. The gender makeup of the item the sales clerk was featured in was changed as well.

#### *Coherence relations coding*

The relationships between the prompts and the continuations given were also examined. Given that certain coherence relations support Goal continuations (next-event mentions) and others support Source continuations (explanations or motivations for the events), it was important to code for and analyze the types of continuations participants provided to consider all possible contributors to referential form. Using Rohde's coding schema<sup>1</sup> (supplied Appendix E) two undergraduate research assistants independently coded each continuation. The seven categories they used were elaboration, explanation, occasion, parallel, result, violated expectations, and background. Two of these continuations (result, occasion) describe events that occur as a result of another event or after it temporally, and thus are more consistent with Goal as

---

<sup>1</sup> We are very grateful to Hannah Rohde for sharing her coding schema with us for this project.

opposed to Source continuations. After coding all the items, the RAs then compared their ratings. On 147 of the 426 total they had coded items differently, so these items were discussed until they had reached an agreement about the appropriate coding.

The continuation codings were divided into two groups for analysis and ease of interpretation- those that described the next events (coded as Occasion and Result) and others. These two groups roughly represented continuations that were consistent with Goal continuations (describing the next event, whether causally connected or not) or Source continuations (describing the possible explanation or motivation for the event).

#### *Statistical modeling*

The data were analyzed following the general analytic approach outlined above. Control models were built first with the relevant control variables List, Character's gender, Mention Other Person, and Connective. List identified whether participants were run on List A or B. The control variable Character's gender referred to the gender makeup of the characters in the item: same or different. The control variable Mention Other Person was a binary variable that coded whether or not participants mentioned the non-designated referent in the same clause, after the referent that had been designated and was coded. What constituted a clause was determined by the clause coding scheme in Appendix B. The control variable Connective was a binary variable that captured whether participants had started their utterance with a connective (*and then, then, so, after that...*) or not. A random intercept for participant was included in the control model.

The main effects models contained the retained control variables (those with *t*-values >1.5 in the control model) as well as the critical predictors of semantic predictability (Goal or Source-continuations), referential predictability (Subject or non-Subject-continuations), gender, type of continuation, and order. Type of continuation was a binary variable that differentiated

between Occasion/Result continuations and all others. Semantic predictability, referential predictability, gender, and type of continuation were centered. A random intercept for participant was included in the main effects model, and the participant by Subject/non-Subject and participant by Goal/Source slopes were included and retained if the model converged and the matrix was positive definite.

The interaction model contained all critical predictors from the main effects model, plus any control variables that were significant in the main effects model, and the random effects structure of the main effects model. Interaction terms were included as well, the four main interaction terms listed previously (Subject\*Goal, Goal\*Gender, Subject\*Gender, and Subject\*Goal\*Gender) as well as the Continuation type \* Goal interaction.

The control variables, critical predictors, and random effects included in the main effects model are shown in Table 1. The variables, random effects, and interaction terms included in the interaction model are shown in Table 2.

## Results

### *Semantic and grammatical role effects on pronoun/zero production*

The first question was whether there was an effect of semantic role on pronoun/zero production. Indeed, in our main effects model we found that participants use more reduced forms when referring to Goals of the prior sentence as opposed to sources. As expected, participants used more pronouns or zeros when referring to Subjects of the prior sentence as compared to non-Subjects (see Figure 4 and Table 1).

The interaction model showed that the Goal effect was qualified by an interaction with continuation type that was trending toward significance  $F(1,375)=3.04, p=0.08$  (see Figure 5).

Critically, however, the main effects of thematic role  $F(1,375)=6.75, p=0.01$  and grammatical role  $F(1,19)=67.00, p<.0001$  remained (see Table 2).

Contrasts revealed the interaction was due to a significant difference in pronoun/zero use for Occasion/Result continuations between Goal and Source items  $t(379)=3.50, p=0.0005$ . No difference in pronoun/zero use was seen for Other continuations between Goal and Source items  $t(379)=0.91, p=0.37$  (see Figure 5).

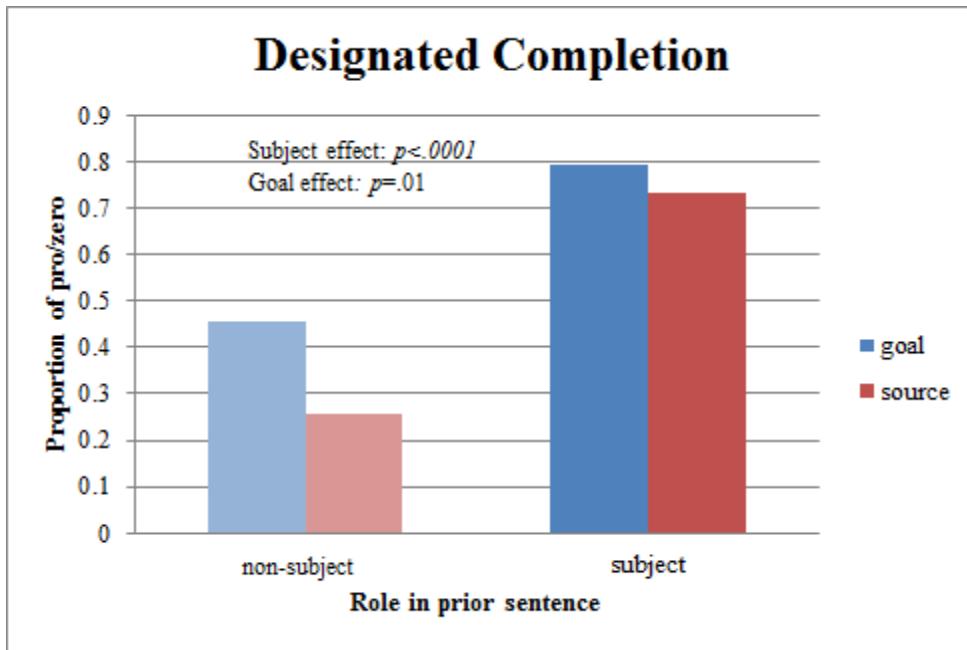


Figure 4. Proportion of pro/zero used by semantic and grammatical roles in prior sentence in Experiment 1

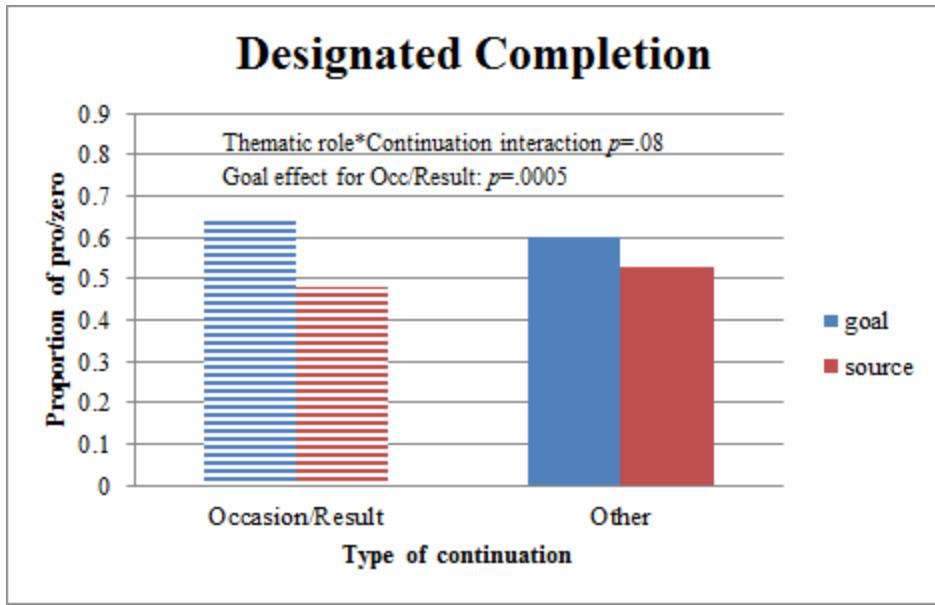


Figure 5. Proportion of pro/zero used by semantic roles and continuation type in prior sentence in Experiment 1

Table 1

Experiment 1 predictor and control variables in the main effects model for referential form choice

	<b>Variable</b>	<b>Estimate</b>	<b>Error</b>	<b>t-value</b>	<b>p-value</b>
<b>Critical Predictors</b>	Goal vs. Source	0.99	0.31	3.17	0.002
	Subject vs. non-Subject	2.85	0.36	7.95	<.0001
	Same gender vs. Different gender	-1.49	0.34	-4.41	<.0001
	Type of continuation (Occ/Result vs other)	0.47	0.32	1.49	0.14
	Order	0.03	0.01	2.87	0.004
<b>Control Variables</b>	List	---	---	---	---
	Mention Other Person	0.82	0.34	2.45	0.015
	Use of connective word	0.82	0.93	0.88	0.38
<b>Random Effects</b>	Participant	*			
	Participant*Subject vs. non-Subject	*			
	Participant*Goal vs. Source	Not positive definite			

Note. T-values for predictor variables indicate their significance. Control variables with t-values >1.5 in the control model were included in the main effects model and their estimates in the main effects model are given here. Dashed lines for control variables indicate the variable was not significant in the control model and thus was not included. Random effects are noted with asterisks if included.

Table 2

Experiment 1 predictor variables, control variables, interaction terms, and random effects in the interaction model for referential form choice

Critical Predictors	Variable	Estimate	Error	t-value	p-value
	Goal vs. Source	0.81	0.31	2.60	0.01
	Subject vs. non-Subject	3.00	0.37	8.19	<.0001
	Type of Continuation (Occ/Result vs other)	0.50	0.32	1.58	0.11
	Same gender vs. Different gender	-1.47	0.34	-4.25	<.0001
	Order	0.03	0.01	2.16	0.03
Interaction terms	Goal *Subject	-0.42	0.58	-0.72	0.47
	Goal*Gender	-0.14	0.65	-0.22	0.82
	Subject*Gender	0.85	0.62	1.37	0.17
	Goal*Continuation type	1.03	0.59	1.74	0.08
	Subject*Gender*Goal	1.17	1.22	0.96	0.34
Control	Mention other person	0.82	0.34	2.41	0.016
	Use of connective word	---	---	---	---
Random Effects	Participant	*			
	Participant by Subject vs. non-Subject	*			
	Participant by Goal vs. Source		Not positive definite		

Note. T-values for predictor variables and interaction terms indicate their significance. Control variables that were significant in the main effects model were included, and T-values indicate their significance in this model. Dashed lines for control variables indicate the variable was not significant in the main effects model and thus was not included. Random effects are noted with asterisks if included.

## Discussion

Experiment 1 found the predicted thematic role effect. Participants produced more pronouns or zeros when referring to prior-Goals than prior-Sources, in contrast with previous findings. They also showed the expected effect of producing more pronouns or zeros for prior Subjects as opposed to prior non-Subjects. Coherence relations were an important modulator of

this effect: when the coherence relation was consistent with the thematic bias (i.e. a prior-Goal in an Occasion/Result continuation), more reduced forms were used than when it was inconsistent.

There was also a main effect of characters' gender. Participants used more pronouns and zeros when the characters were of different genders than when they were the same. This could have resulted from a couple factors. First, this effect could have resulted from ambiguity-avoidance considerations. Use of a pronoun or zero in the different-gender condition would not be ambiguous, but would be in the same-gender condition. Another possibility is that it was the result of semantic competition. Characters of the same gender share more semantic characteristics than characters of different genders, perhaps leading to competition and thus reduced accessibility.

The implications of these findings for the role of predictability in choice of referring expression will be discussed later, after more specific evidence about the role of predictability and its time course. However, the important finding of Experiment 1 is that thematic roles do matter in choice of referring expression. The question then is why we found this effect, when previous studies have failed to. While similar in important ways, our story-continuation task differed from previous designs. Potential important differences in methodology may have allowed us to detect the thematic role effect where it was previously not found.

## **CHAPTER 3: VARIATIONS ON METHODOLOGY**

Experiment 1 found an effect of semantic role predictability on pronoun/zero production: participants used more reduced forms to refer to Goals as opposed to Sources of the prior sentence. Experiment 1 differed from previous studies in several ways. The stimuli used were composed of Goal-Source verbs, as opposed to Implicit Causality verbs. Unlike in previous studies, participants in Experiment 1 who did not provide any variation in their responses were excluded. The items in Experiment 1 were also not totally independent of each other, as they described the same participants and used repeated mention of items. To ensure equal numbers of items across the conditions, participants in Experiment 1 were also instructed which character to refer to (as has been done previously, e.g. Fukumura & van Gompel, 2010). Slight variations to Experiment 1 were made to create Experiments 2 and 3, to determine what components of the design were critical in order for the effect to emerge.

### **EXPERIMENT 2: ONLINE RENAMED ITEMS DESIGNATED SENTENCE COMPLETION**

#### *Motivation*

Experiment 2 was conducted to determine whether the semantic predictability effect found in Experiment 1 was dependent upon the experimental items being related to one another. It may have been the case that the semantic predictability effect was due to the fact that participants were able to build a mental model of the events as a whole, since the items were related. A richer representation of the overall framework of the items may have allowed participants to dedicate more resources to utilizing predictability information, as they did not have to construct new representations of characters with every item. Experiment 2 eliminated

repeated mention of people and items while maintaining the same experimental controls on grammatical role, thematic role, and gender.

## Method

### *Participants*

57 participants completed the task on Amazon Mechanical Turk, all for a monetary reward. They needed to be native English speakers with a HIT approval rate greater than or equal to 98%, with at least 5000 approved HITs. As in Experiment 1, participants also needed to use two pronouns or zeros and two names in their responses in order to be included. 37 participants were excluded, leaving a total of 20 participants for whom data was analyzed. Of the 37 who were not analyzed, five were excluded for doing any earlier version of the experiment<sup>2</sup>, three were excluded for providing meaningless continuations, 25 were excluded for using fewer than two pronouns or zeros, and four were excluded for using fewer than two names.

### *Materials and Design*

The items for Experiment 2 were created by amending the items used in Experiment 1, and are in Appendix C. Critically, the control of the stimuli features remained the same, such that the items were balanced by grammatical role and gender. All the verbs remained the same. The names/occupations and items were changed, however, such that no name/occupation or item was repeated. If the original item mentioned a character by name, that name was replaced with another common name of the same gender (the names were selected from a list of the most popular male and female names in 1958). If the character was mentioned by occupation in the original item (butler, nurse), then that occupation was replaced with a common occupation. All attempts were made to replace occupations with other occupations that typically are gender-specific, to preserve the same and different gender makeup of the stimuli. Objects that had been

---

<sup>2</sup> The order in which the experiments are reported here is different than the order in which they were run.

mentioned in the original items were replaced with other common objects such that none were repeated. Experiment 2 stimuli can be found in the appendix. A sample item from Experiment 2 is given in Figure 6.

Michael received a painting of the two of them from Mary.

|

Figure 6. Sample trial from Experiment 2

#### *Procedure*

The procedure was identical to that of Experiment 1; participants were presented with prompt sentences and were asked to provide a plausible continuation sentence about the character that was underlined.

#### *Analysis*

##### *Response Coding*

The inclusion criteria were identical to those of Experiment 1. Given the very high consistency of ratings between the original coder and the re-coders for Experiment 1, no double coding was performed for this experiment. The same coding criteria applied for this Experiment (e.g., responses were coded for who was referred to and how they were referred within the first clause of the response). Fifty-eight items were excluded from the final analysis, leaving 422 items. The items were split evenly among Subject and non-Subject conditions (211 items each) and fairly evenly among Goals (222 items) and Sources (200 items).

Seventeen items were excluded for being about inanimate referents, one was excluded for being a plural Subject, five were excluded for being possessive, one was excluded as the intended referent was unclear, one was excluded for using ‘who’ as the Subject, and 33 were excluded as the wrong character was referred to. Responses were coded for use of connective

words such as then, and, etc., and for whether the non-designated character was mentioned in the same clause as the coded referent.

#### *Gender makeup of characters*

The gender makeup of the characters was maintained in the experimental coding, as participants interpreted the gender of the characters as had been intended.

#### *Statistical Modeling*

The data were analyzed in an identical manner to Experiment 1. A control model was first built, then a main effects model, then an interaction model. Significant control variables were retained for the main effects model, and significant control variables in the main effects model were retained for the interaction model, along with the critical predictors and random effects. Participant was included as a random intercept in all the models and random slopes of participant by Subject/non-Subject and participant by Goal/Source were included in the main effects models and interaction model if the model converged and was positive definite.

### Results

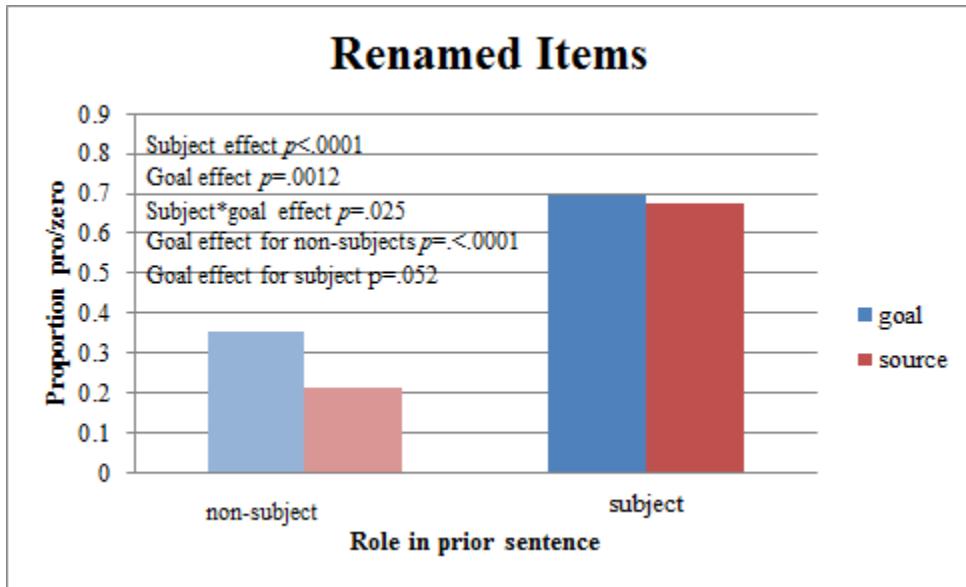
#### *Semantic and grammatical role effects on pronoun/zero production*

The main effects model, shown in Table 3, revealed that participants used more pronouns/zeros to refer to Goals of the prior sentence as compared to Sources. They also used more reduced forms to refer to Subject of the prior sentence as opposed to non-Subject.

In the model including interactions (see Table 4) these effects were retained: there was a main effect of Goal  $F(1,375)=10.70, p=.0012$ , and a main effect of Subject  $F(1,19)=66.66, p<.0001$ , yet these effects were qualified by interactions.

There was a Goal by Subject interaction  $F(1,375)=-5.07, p=0.025$ (see Figure 7), and a Goal by gender interaction  $F(1,375)=5.06, p=0.025$ (see Figure 8). Contrasts in the model

suggested that the interaction between Subject and Goal was due to a stronger effect in the non-Subject condition  $t(377)=4.01, p<.0001$  than in the Subject condition  $t(377)=1.95, p=.052$ .



*Figure 7.* Proportion of pro/zero used by semantic and grammatical role in prior sentence in Experiment 2

The Goal by gender interaction, as seen in Figure 8, was revealed by contrasts to be due to the difference in pronominalization rates between Goal and Source continuations in the same gender condition  $t(378)=3.50, p=.0005$ . There was no difference in the different gender condition  $t(378)=0.94 p=.35$ .

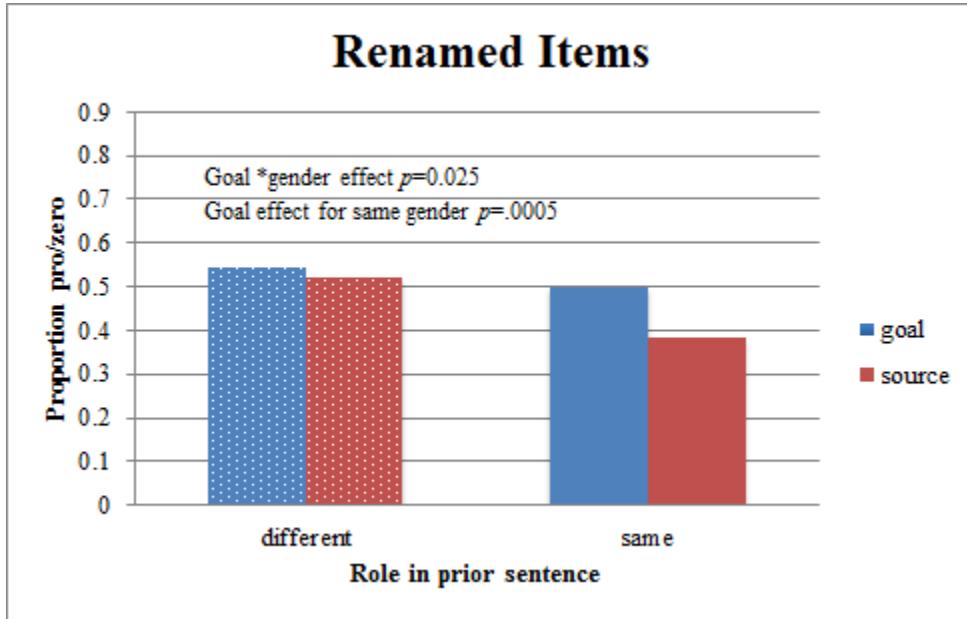


Figure 8. Proportion of pro/zero used by semantic roles and gender makeup of characters in prior sentence in Experiment 2

Table 3

Experiment 2 predictor and control variables in the main effects model for referential form choice

Critical Predictors	Variable	Estimate	Error	t-value	p-value
	Goal vs. Source	0.93	0.31	2.99	0.003
	Subject vs. non-Subject	3.06	0.38	8.08	<.0001
	Same gender vs. Different gender	-0.78	0.28	-2.74	0.006
	Order	0.044	0.01	4.36	<.0001
Control Variables	List	---	---	---	---
	Mention Other Person	---	---	---	---
	Use of connective word	---	---	---	---
Random Effects	Participant	*			
	Participant by Subject vs. non-Subject	*			
	Participant by Goal vs. Source	Not positive definite			

*Note.* T-values for predictor variables indicate their significance. Control variables with t-values >1.5 in the control model were included in the main effects model and their values in the main effects model are given here. Dashed lines for control variables indicate the variable was not significant in the control model and thus was not included. Random effects are noted with asterisks if included.

Table 4

Experiment 2 predictor variables, interaction terms, and random effects in the interaction model for referential form choice.

Critical Predictors	Variable	Estimate	Error	t-value	p-value
	Goal vs. Source	1.07	0.33	3.27	0.0012
	Subject vs. non-Subject	3.24	0.40	8.16	<.0001
	Same gender vs. Different gender	-0.92	0.30	-3.09	0.002
	Order	0.05	0.01	4.58	<.0001
Interaction terms	Goal *Subject	-1.35	0.60	-2.25	0.025
	Goal*Gender	1.35	0.60	2.25	0.025
	Subject*Gender	0.68	0.58	1.14	0.25
	Subject*Gender*Goal	-1.91	1.19	-1.61	0.11
Random Effects	Participant	*			
	Participant by Subject vs. non-Subject	*			
	Participant by Goal vs. Source	Not positive definite			

*Note.* T-values for predictor variables and interaction terms indicate their significance. No control variables were used in the main effects model and thus none were used here. Random effects are noted with asterisks if included.

### Discussion

As found in Experiment 1, participants referred to prior Goals with more pronouns or zeros as compared to prior Sources. As expected, participants continued to prefer using reduced forms when referring to Subjects of the prior sentence as opposed to non-Subjects. As the experimental items were no longer related to one another, this effect could not be due to any coherent, overarching model of the events. There was also a Subject by Goal interaction, such that participants referred to non-Subject continuations with more pronouns/zeros when they were Goals of the prior sentence as opposed to Sources. This effect was not seen in the Subject continuations, but this may be due to participants' high overall pronoun/zero production in this condition.

There was also a main effect of characters' gender, as seen in Experiment 1. Participants used more reduced forms to refer to characters of different genders than same-gender characters. There was an interaction of gender by Goal. In the same gender condition, participants used significantly more pronouns and zeros for Goals than Sources. When a pronoun would be ambiguous, information about its predictability had a significant effect on what form was selected. When a pronoun would not be ambiguous (such as in the different-gender condition), thematic role predictability information did not affect the form.

### EXPERIMENT 3: ONLINE FREE SENTENCE COMPLETION

#### *Motivation*

Experiment 3 was conducted to achieve two aims. The first was to determine whether the semantic bias effect was strong enough to be found without the experimental control of designated completions, which ensured equal numbers of items in each condition. A second motivation was to provide independent conformation for the underlying assumption that Goals are more predictable than Sources. One way to evaluate this assumption is to examine the likelihood that the speaker mentions the Goal or Source in their continuations. The previous experiments could not examine this question, as participants had been instructed who to refer to.

#### Method

##### *Participants*

63 participants completed the task, 28 for course credit and 35 for a monetary reward. 23 participants were excluded, leaving a total of 40 participants for whom data was analyzed. Of the 23 who were not analyzed, one was excluded for having incomplete data, four for using fewer than two proper names, 15 for using fewer than two pronouns or zeros, two for providing meaningless continuations, and one for completing an earlier version of the study. The inclusion

criteria for the participants who received course credit (28) were that they must be native English speakers, have normal or correct-to-normal vision, and couldn't have participated in a similar experiment in the lab. As in Experiment 2, the participants who completed the experiment for payment (35) via Amazon Mechanical Turk needed to be native English speakers with a HIT approval rate greater than or equal to 98%, with at least 5000 approved HITs.

#### *Materials and Design*

The experimental trials in this experiment consisted of the sentences (24 critical items and 29 fillers) described earlier and used in Experiment 1. Participants were recruited through Amazon Mechanical Turk or through SONA. In either case, the sentences were presented to participants with a computerized survey through Qualtrics. In this free-completion experiment participants were instructed to provide a plausible continuation about any of the characters mentioned in the first sentence. Participants were presented with one of the two lists created, allowing them to see each item in one of the two conditions, but both conditions across items. The stimuli for Experiment 1 can be found in Appendix D. An example item from Experiment 3 is shown in Figure 9:

Sir Barnes received a painting of the two of them from Lady Mannerly.

*Figure 9. Sample trial from Experiment 3*

#### *Procedure*

Participants were presented with the first sentence of each pair, as described above. Participants were asked to provide a plausible continuation about either of the characters mentioned in the sentence. Participants provided typed responses via computer.

#### *Analysis*

#### *Response coding*

The Subjects of the first clauses of participants' responses were coded, using the same coding schema as in Experiments 1 and 2. These responses were coded for whom the participant had chosen to speak about and how they referred to that person. That is, responses were coded for choice of referring expression (pronoun/zero or proper name) and role of referent in the prior sentence (Subject or non-Subject and Goal or Source). Items were excluded if participants referred to more than one person at once (e.g., *Then they put the groceries away*), they did not refer to someone who was in the prompt sentence, they referred to some inanimate object as the Subject, or they referred to someone's possession or body part as the Subject (e.g., *Sir Barnes' back was sore*). Participants who did not use at least two pronouns or zeros or at least two names were excluded from the analysis altogether.

The author did an initial coding of all the items for the first 20 participants, and then the items were re-coded by one of two undergraduate research assistants who were blind to the author's coding. Of the 480 total items there were 20 mismatches between the original and re-coding, or 4.2% of the data. Of those 20, 11 were cases in which there was a disagreement about who the intended referent was, or one of the coders thought it was unclear. Of these 11, seven were excluded because it was impossible to definitely decide whom the intended referent was or because once we agreed upon the intended referent it met exclusion criteria (e.g., the referent was not in the prior sentence). The remaining four were included, as the author and the second coder came to a consensus about who was being referred to. Nine of the 20 mismatches were cases in which with second coder miscoded or neglected to code the item. In these nine cases the author and the second coder conferred and agreed that the original coding was correct.

Of the 960 items for all 40 participants, 109 were excluded from the final analysis, leaving 851 items. Sixty-four items were excluded for being about inanimate referents, five were

excluded for referring to a character not in the prior sentence, 17 were excluded for including a plural character, four were excluded for being possessive Subjects, eight were excluded as the referent was unclear, seven were excluded because the first clause in the sentence did not contain a clear Subject (e.g., ‘*taking a deep breath, ...*’) one was excluded for using ‘*who*’ to refer to the character, and three were excluded because the Subject was part of a quotation (e.g., “*Do it correctly, girl*”, she warned).

Whether participants used connectives at the beginnings of their continuations and whether they mentioned another character in the main clause were coded, as in Experiments 1 and 2.

#### *Gender makeup of characters*

The chef, intended to be female for the purposes of balancing the gender makeup of the characters, was interpreted as male by most of the participants. The coding of gender for this data was changed to reflect this. Another ambiguously gendered character, the sales clerk, was only referred to three times (*the sales clerk, the sales clerk, he*), so this character maintained its original gender coding as male.

#### *Statistical modeling*

The same analytical approach was used as for Experiments 1 and 2. To determine whether participants preferred to refer to Goals over Sources and Subjects over non-Subjects, two random-intercept only models were run with random intercepts for participants and choice of referent as the outcome. This allowed us to determine whether participants had a significant preference to select either the Goal or the Source as the more likely character to be continued. The same procedure was done with Subject or non-Subject chosen as the outcome measure and a

random intercept for person, to determine whether participants had a preference for either Subject or non-Subject to be continued.

In models that are run with no predictors, the intercept can be interpreted as the estimated log odds of the outcome occurring. In this case, that outcome was whether they referred to the Goal, in one model, or the Subject, in another model. These odds can be transformed into a probability with the formula  $p = \text{exponential function (intercept)} / (1 + \text{exponential function(intercept)})$ .

## Results

### *Semantic and grammatical role effects on pronoun/zero production*

As expected, we saw strong evidence for the Subject-bias in pronoun use in the main model. Participants used more reduced forms when referring to the Subject of the prior clause as compared to the non-Subject (see Table 5 for model details). The critical question was whether thematic roles would significantly affect pronoun/zero use. Although the graph (Figure 10) showed the expected numerical pattern, this was not a significant effect. The interaction model suggested that the Subject main effect was qualified by an interaction with characters' gender,  $F(1,764)=6.53, p=0.01$ , but there was still a main effect of Subject,  $F(1,39)=161.37, p<.0001$ . See Table 6 for model details and Figure 11 for the interaction. There was also a main effect of character's gender,  $F(1,764)=43.14, p<.0001$ .

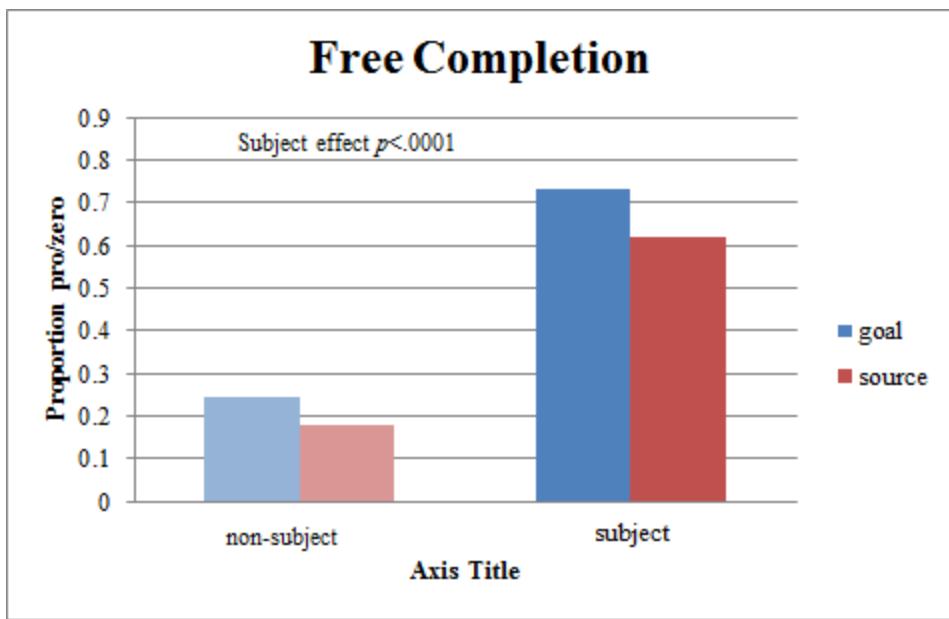


Figure 10. Proportion of pro/zeros used by semantic and grammatical roles in prior sentence in Experiment 3

Contrasts included in the model revealed significant Subject effects in both the different gender condition,  $t(767)=9.69, p < .0001$ , and in the same gender condition,  $t(767)=10.24, p < .0001$ , suggesting the interaction was due to the slightly stronger effect in the same gender condition.

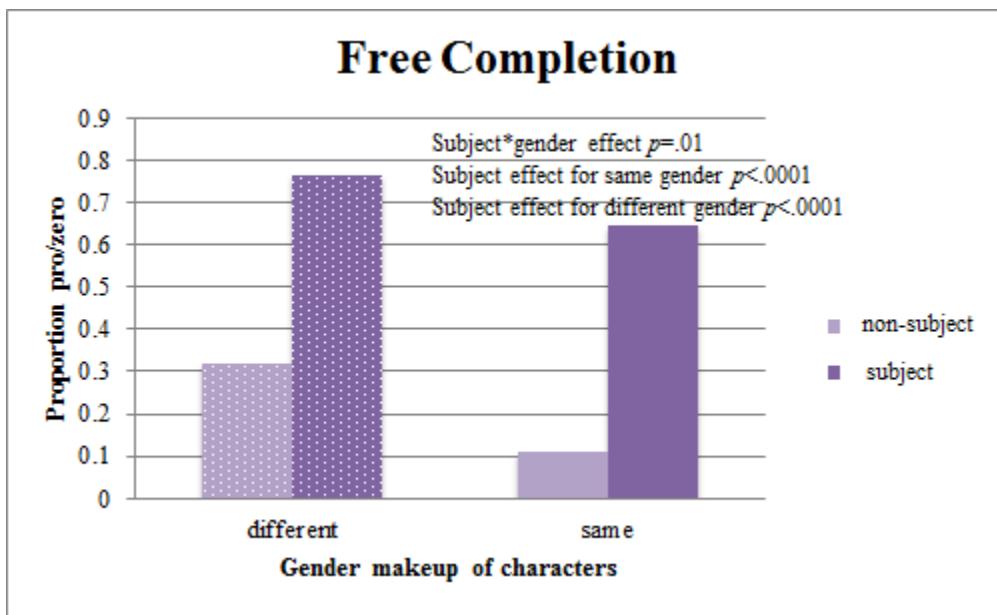


Figure 11: Proportion of pro/zero used by grammatical roles and makeup of characters' gender in prior sentence in Experiment 3

Table 5

Experiment 3 predictor and control variables in the main effects model for referential form choice

	<b>Variable</b>	<b>Estimate</b>	<b>Error</b>	<b>t-value</b>	<b>p-value</b>
<b>Critical Predictors</b>	Goal vs. Source	0.23	0.26	0.88	0.38
	Subject vs. non-Subject	3.21	0.25	12.83	<.0001
	Same gender vs. Different gender	-1.46	0.22	-6.46	<.0001
	Order	0.02	0.007	2.34	0.02
<b>Control Variables</b>	List	---	---	---	---
	Mention Other Person	---	---	---	---
	Use of connective word	---	---	---	---
<b>Random Effects</b>	Participant	*			
	Participant by Subject vs. non-Subject	*			
	Participant by Goal vs. Source		Not positive definite		

*Note.* T-values for predictor variables indicate their significance. Control variables with *t*-values >1.5 in the control model were included in the main effects model and their values in the main effects model are given here. Dashed lines for control variables indicate the variable was not significant in the control model and thus was not included. Random effects are noted with asterisks if included.

Table 6

Experiment 3 predictor variables, interaction terms, and random effects in the interaction model for referential form choice

Critical Predictors	Variable	Estimate	Error	t-value	p-value
	Goal vs. Source	0.24	0.29	0.84	0.40
	Subject vs. non-Subject	3.26	0.26	12.70	<.0001
	Same gender vs. Different gender	-1.58	0.24	-6.57	<.0001
	Order	0.019	0.007	2.49	0.01
Interaction terms	Goal *Subject	0.29	0.56	0.53	0.60
	Goal*Gender	0.04	0.60	0.07	0.94
	Subject*Gender	1.11	0.43	2.55	0.01
	Subject*Gender*Goal	-0.22	1.19	-0.19	0.85
Random Effects	Participant	*			
	Participant by Subject vs. non-Subject	*			
	Participant by Goal vs. Source	Not positive definite			

Note. T-values for predictor variables and interaction terms indicate their significance. No control variables were used in the main effects model and thus none were used in this model. Random effects are noted with asterisks if included.

#### *Counts of references in each condition*

Confirming the next-mention bias for Goals, participants showed an overwhelming preference to provide continuations about the Goal as opposed to the Source. As shown in Figure 12, participants provided 362 continuations about the Goal when it was the non-Subject of the prompt sentence and 335 continuations about the Goal when it was the Subject of the prompt sentence. Participants provided 83 continuations about the Source when it was the non-Subject of the prior sentence and 71 when it had been the Subject. The random intercept model of the count of references to the Goal, run with no predictors, confirmed the significance of this effect. The likelihood of an average person choosing to speak about the Goal was 83%, which was significantly higher than 50%  $t(39)=11.92, p<.0001$ . The intercept estimate in this model was

$1.57, p=\text{exponential function}(1.57)/1+\text{exponential function}(1.57)$ . Participants did not show a preference to talk about the Subject as opposed to the non-Subject. The probability of talking about the subject was 48%. The intercept estimate was  $-0.09, p=\exp(-0.09)/1+\exp(-0.09)$ .

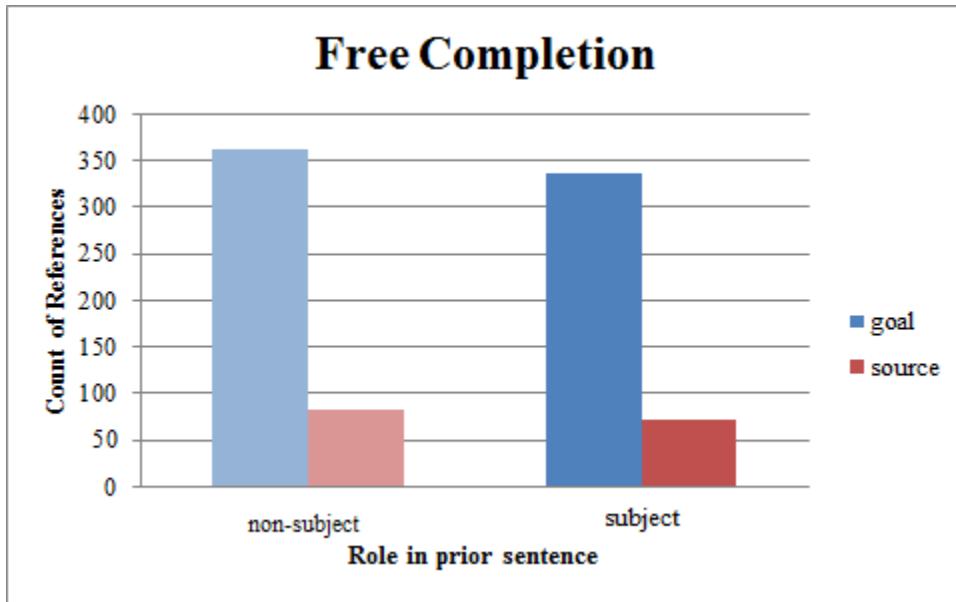


Figure 12. Count of total references by grammatical and semantic roles in the prior sentence in Experiment 3

### Discussion

Experiment 3 found a significant effect for choice of referring expression; participants preferred using pronouns/zeros to talk about characters that had been the Subject in the prompt sentence as compared to the non-Subject. There was a significant grammatical role by gender interaction, but contrasts revealed this was due only to a slightly larger effect for the same gender items.

No effect of semantic role as Goal was found for choice of referring expression. The lack of a semantic role effect may have been due to limited power, given the numerical trend in this direction. There were very few references for the less predictable referents (the Sources), compared to a vastly greater number of Goal continuations.

Experiment 3 also found the predicted pattern of speakers preferring to provide continuations about the Goal of the prompt sentence far more often than the Source. Speakers preferred to begin their continuations with the Goal of the prior sentence as opposed to the Source. Speakers did not show a preference to begin their continuations about the Subjects as compared to the non-Subject.

First, this study provided confirmatory evidence that Goal referents are more predictable than Sources (as measured by choice of referent). It also showed that Subject referents were not found to be more predictable than non-Subject referents, as they were chosen as the continuation an equal amount of time. Although Goals were found to be much more predictable than Sources, they were not referred to with more pronouns or zeros (although this may have been due to an issue of power).

### DISCUSSION OF EXPERIMENTS 1-3

The initial question this set of studies was asking was whether predictability mattered for choice of referring expressions. The answer is yes: the predictability of thematic roles influences referential expressions. Participants used more reduced referring expressions (pronouns or zeros) to refer to Goals as compared to Sources.

The results of Experiment 3 highlight the fact that the semantic bias effect is small, especially in comparison to the grammatical role effect, which was robust in Experiments 1-3. The findings of Experiment 3, in isolation, are consistent with earlier claims that only grammatical role influences pronoun/zero production. However, within the larger set including Experiments 1 and 2, there is strong evidence that semantic role has an effect on reference production. Participants used significantly more reduced forms to refer to Goals versus Sources in the two earlier experiments. In all three experiments they showed the predicted effect of using

pronouns or zeros to refer to Subjects at a higher rate than for non-Subjects. Taken together, these three experiments indicate that the semantic bias effect, while small, does exist when the right methods are used to find it. While generally supporting the same main effects, the differences in methodology between the three experiments may have had implications for the precise findings, such as what interactions were found. Given the inconsistency of the interactions between the three experiments on the rate of pronoun/zero use, it would be difficult to state what the true interactions are without a more systematic comparison of different variations.

Experiment 2 indicated that an overarching discourse context, between items, is not required to find the semantic bias effect on referring expressions. This suggests that, at least within the sentence continuation paradigm, participants are capable of rapidly constructing mental models for each item that allow them to form coherence relations between the stimuli and their constructed continuation. Whether participants could construct these models for unrelated items within different task constraints, such as under time pressure or with the added pressure of an addressee, remains to be seen.

Experiment 3 suggested the critical importance of equal numbers of items in each condition to determine the existence of this effect. While numerically it appears that participants used more pronouns and zeros for Goals as opposed to Sources, the effect was not significant. When participants are not forced to provide continuations about Sources they greatly prefer to talk about Goals.

Another important component of these experiments is the exclusion criteria applied to linguistic form variation. Participants needed to provide some variation in their referring expressions, and only participants who used two names and two pronouns or zeros were

included. Previous studies included participants regardless of linguistic form variation. There is some evidence that the prior lack of finding a semantic role effect may have been due to not excluding such participants. An analysis of the first 10 participants on each list, regardless of form variation, in the experiment most similar to those previously conducted (the renamed items study) found no effect of Goal-hood on pronoun/zero use, but a significant effect of Subject-hood.

Given the relatively small size of the semantic bias effect, these experiments suggest that certain conditions must be in place to find it: (1) sufficient and equal numbers of items in each condition, and (2) variation in the referring expressions given by participants. An overarching discourse context between items is not necessary. Whether these conditions are sufficient to find the effect with any class of verbs, or whether they are only sufficient to find it only with Goal-Source verbs is the topic of further ongoing study.

## **CHAPTER 4: FURTHER ANALYSIS OF PREDICTABILITY**

Experiments 1-3 suggest that semantic roles do affect referential expressions, and that participants used more reduced forms to refer to Goals as opposed to Sources. The next question was: by what mechanism does this occur? Experiment 4, an in-person event-retelling study, was conducted in order to measure and test factors such as latency, which could inform potential mechanisms.

Experiment 4 also served as a test of the manipulation: we wanted to be sure that the effects found in the computerized, written continuation experiments were consistent with the results of a spoken, in-person design. If the results of Experiments 1-3 could not be duplicated in an in-person design, that would suggest that the effect depended on the methods used in the computerized story-continuation tasks, such as the relatively unlimited time constraints on planning or the written response modality. In that case, an examination of the underlying cognitive mechanism would be premature.

### **EXPERIMENT 4: IN PERSON DESIGNATED EVENT RETELLING**

#### *Motivation*

Experiment 4 was conducted to determine whether the semantic predictability effect found in the online sentence completion experiments could be replicated with the same stimuli in a more interactive setting. Many of the features of this task differed from the sentence continuation design, as participants were speaking as opposed to writing, interacting with a confederate, and describing pictures they were familiar with instead of constructing a plausible

continuation. Critically, the linguistic features of the stimuli they were presented with remained the same.

### Method

#### *Participants*

32 undergraduates completed the task for class credit. 12 participants were excluded, leaving a total of 20 participants for whom data was analyzed. 10 participants were excluded for using fewer than two pronouns or zeros and two were excluded for being non-native English speakers. As for the previous experiments, participants needed to use two pronouns or zeros and two names in their responses to be included.

#### *Materials and Design*

Participants viewed pairs of pictures that were depictions of the sentence pairs described above. Participants heard a description of the first picture in each pair, produced by a lab confederate, and then provided their own description of the second. The stimuli were divided across two lists as in the other experiments, such that each participant saw the same pictures, but heard one of two versions of the critical prompt sentences. Experiment 4 stimuli pictures can be found on the supplementary materials website<sup>3</sup>. A sample item from Experiment 4 is given in Figure 13.

---

<sup>3</sup> jaapstimuli.web.unc.edu

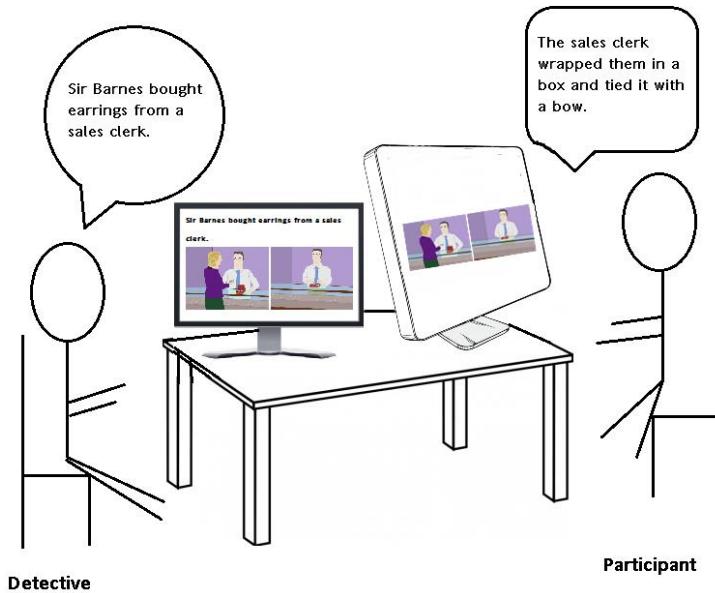


*Figure 13.* Sample trial from Experiment 4

#### *Procedure*

Participants were brought into the lab and seated at a computer. Participants were consented and completed an optional participant questionnaire, and then were shown a narrated slideshow (transcript in Appendix F). The slideshow told them that they were a tabloid photographer, and described the family they had been secretly taking photographs of. It then told them that a murder occurred while they were at the house, and they were going to review the photographs they had taken to help a detective solve the crime. The participants were introduced to the characters in the pictures, and then were shown all their pictures, in order. After the participants had been shown the pictures they completed a sample item with the lab confederate, who explained that the detective, who would arrive shortly, would describe the first picture in the pair. After that, the participant should say what happened next, using the second picture as a guide.

The detective then entered the room and asked the participant to recount who the family was they had been photographing. Then the audio recorder was turned on and the detective sat down at her own computer, whose back was to the back of the participant's computer. The participant's monitor was large enough that the participant and detective could not see one another. The detective and participant then began looking at the pairs of pictures together. The detective would describe the first picture using a script (identical to the prompt sentences given in Experiments 1 and 3), and the confederate would then say what happened next, by referring to the second picture displayed on her computer. Both pictures in the pair appeared at once on the screen, to encourage participants' conception of them as a coherent set. After the participant described the second picture the detective would then advance the pairs of pictures on both computers simultaneously. A depiction of this set-up can be seen in Figure 14.



*Figure 14.* Experimental set-up from Experiment 4

When the detective and participant had described all the events, the detective then asked the participant who had been murdered, who had committed the crime and with what weapon, and why. The detective then told the participant they could both come out of character. The detective proceeded with further debriefing questions about the participants' familiarity with the Clue game. The post-experiment questionnaire can be found in Appendix G.

### *Analysis*

#### *Response coding*

The inclusion criteria were identical to those of the previous experiments. Participants needed to refer to the character pictured in the second picture of each pair for the item to be included. Given the very high consistency of ratings between the original coder and the re-coders for previous experiments, no double-coding was performed for this experiment.

56 items were excluded from the final analysis, leaving 424 items. The items were evenly divided among Subjects (203 items) and non-Subjects (221 items) and Goals (218 items) and Sources (206 items). 24 items were excluded for being about non-human referents, 24 were excluded as the wrong character was referred to, one was excluded for being a plural Subject, one was excluded as the slide was advanced too soon to the next trial, one was excluded for using 'who' as the Subject, and five were excluded because of other mechanical issues (two pictures were advanced instead of one; the picture was advanced too soon, etc). The data were analyzed in an identical manner to Experiments 1, 2 and 3.

Responses were also coded for use of a connective (after, afterwards, and, and then, next, now, then, after that), which was included as a control variable in the model. Use of a connective was hypothesized to indicate increased use of the discourse and conceptualization of the two events as a unit. Whether participants mentioned the other character in their continuation was

also coded as a control variable, as in the previous three experiments. Any disfluencies at the onset of responses were also coded.

#### *Audio data coding*

The audio data were analyzed with Praat to measure latency to begin speaking. A primary undergraduate research assistant coded all the included items, making note of three time points: the end of the beep, which signaled the presentation of the pictures for a particular trial, the end of the detective's speech (the description of the first picture), and the onset of the participant's speech (describing the second picture). A second research assistant double coded all the items as a check of reliability. Two latencies were obtained for each research assistant from these three time points, one which measured the time that elapsed between the beep and the participant speaking (hereon latency after beep) and the time that elapsed between the end of the detective's speech and the participant speaking (hereon latency after detective). Latencies after the beep were obtained for each item by subtracting the beep time point from the beginning of the participant's speech time point. Latencies after the detective were obtained for each item by subtracting the time point at the end of the detective's speech from the beginning of the participant's speech time point.

These latencies were then compared between the two coders. Latencies that were more than 10% different from one another were analyzed by the author (ER). The author either selected the coding she thought was correct, or if she thought both were incorrect, she supplied the correct latency.

For the rest of the cases in which the latencies did not differ more than 10% between the two research assistants, the latencies of the first research assistant were used.

Twenty-three of the latencies after the beep differed by more than 10% from one another (making up 5% of the total items). Of these 23, the author found that 8 were correctly coded by the first research assistant, 14 were correctly coded by the second research assistant, and in one case she agreed with neither coder, and supplied the correct latency.

100 of the latencies after detective differed by more than 10% from one another, which comprised 23% of the total items. Many of these cases were ones in which the offset of the detective's speech was subtle, or one of the coders misapplied the rule concerning fillers (um was to be included in the latency period, rather than as the onset of the speech). In 30 of these 100 cases the author agreed with the coding of the first research assistant, in 33 she agreed with the coding of the second, and in 37 cases she disagreed with both and supplied her own latency measurement. The latency after detective will be used for the latency analysis.

#### *Gender makeup of characters*

As participants could view the depictions of characters and their actions, we could ensure that the genders of the characters were interpreted as had been intended.

#### *Statistical modeling*

As in Experiments 1, 2, and 3, whether participants had referred to characters with a pro/zero or name was analyzed with a 3-model process, using SAS proc glimmix.

The same 3-model approach was taken to analyze participants' use of connective words (*and, then, after that*, etc). This approach was also taken to analyze the latency data with SAS proc mixed, because the outcome is continuous. Several of the control variables of latency were correlated with each other (correlations shown in Table 13). Each of these variables was tested individually, in isolation of the others, in the model, and if it was significant it was retained.

The same variables that were controls in previous experiments (List, Order, Character's gender, Mention other person, Connective) were included in these control models and their definitions are identical to previous descriptions. Additional control variables were also included in these models. The variable Referent on right in picture refers to whether the continuation referent was on the right side of the first picture of each pair. Relatedness Z-score is a variable from Experiment 5, and will be discussed in more detail later. Briefly, it is a rating of how related the two events in each pair were judged to be, by a separate set of participants. Likelihood of mention is a variable from Experiment 6, and will also be discussed in greater detail later. It is a measure of how likely the designated referent on that item was judged to be, by a different set of participants. Word count was a measure of how many words the participant used in that particular utterance. Ease of verb was computed, for each participant and each item, by examining the verbs each participant used to describe the action pictured. For example if 18/20 participants described an action as shooting, those participants got a score of 18/20 for that item. The other two participants who describe it as something else (*loading the gun*, for example), got a score of 2/20 for that item. The variable disfluency was coded as a 1 if participants were disfluent at the beginning of their utterance and 0 if they were not.

## Results

### *Semantic and grammatical role effects on pronoun/zero production*

The critical effect of thematic role was found in the main effects model of Experiment 4, as shown in Figure 15 and Table 7: participants used more pronouns/zeros when referring to Goals of the prior sentence as compared to Sources,  $F(1,19)=7.96, p=0.01$ . As was expected, they also used more pronouns and zeros when referring to Subjects of the prior sentence as compared to non-Subjects,  $F(1,19)=19.92, p=0.0003$ . The Goal by Subject interaction ( $p=.12$ )

approached the significance cut-off for examination, but the main effects remained significant in the interaction model (see Table 8 for model details). Visual examination of Figure 15 suggests the interaction is due to the stronger Goal effect in the Subject condition.

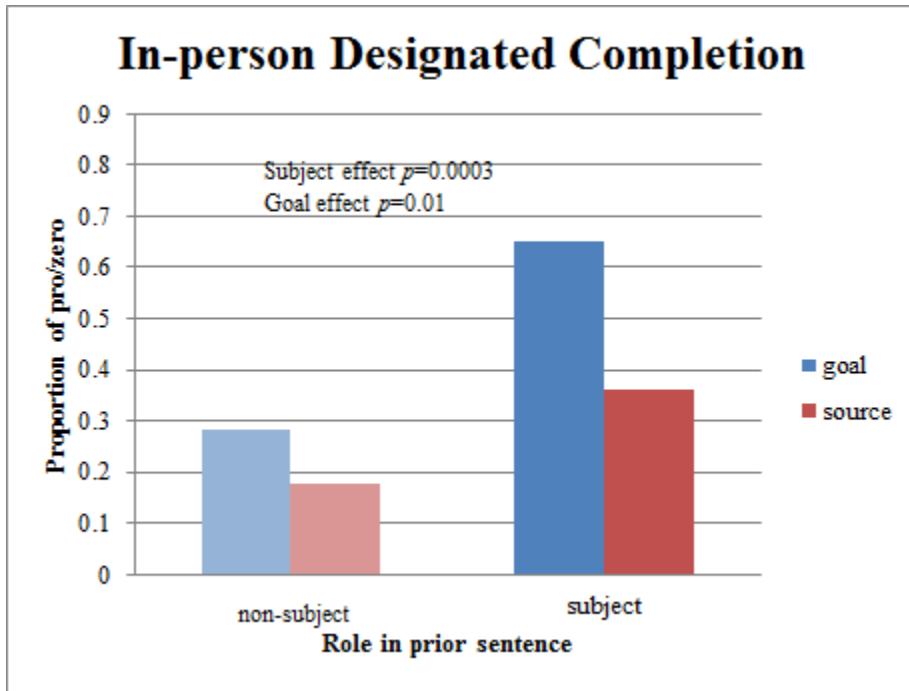


Figure 15. Proportion of pro/zero used by semantic and grammatical roles in prior sentence in Experiment 4

Table 7

Experiment 4 predictor and control variables in the main effects model for referential form choice

Critical Predictors	Variable	Estimate	Error	t-value	p-value
	Goal vs. Source	1.03	0.36	2.82	0.01
	Subject vs. non-Subject	1.44	0.32	4.46	0.0003
	Same gender vs. Different gender	-0.30	0.25	-1.24	0.22
	Order	-0.017	0.008	-2.00	-0.05
Control Variables	List	---	---	---	---
	Mention Other Person	---	---	---	---
	Use of connective word	0.89	0.30	2.96	0.003
	Referent on right in picture	---	---	---	---
	Relatedness	---	---	---	---
	Likelihood	0.04	0.62	0.06	0.95
	Count of words	---	---	---	---
	Ease	---	---	---	---
	Disfluency	---	---	---	---
	Latency to begin speaking	---	---	---	---
Random Effects	Participant	*			
	Participant by Subject vs. non-Subject	*			
	Participant by Goal vs. Source	*			

*Note.* T-values for predictor variables indicate their significance. Control variables with t-values >1.5 in the control model were included in the main effects model and their values in the main effects model are given here. Dashed lines for control variables indicate the variable was not significant in the control model and thus was not included. Random effects are noted with asterisks if included.

Table 8

Experiment 4 predictor variables, control variables, interaction terms, and random effects in the interaction model for referential form choice

Critical Predictors	Variable	Estimate	Error	t-value	p-value
Critical Predictors	Goal vs. Source	0.97	0.29	3.32	0.0036
	Subject vs. non-Subject	1.40	0.32	4.38	0.0003
	Same gender vs. Different gender	-0.30	0.24	-1.22	0.22
	Order	-0.017	0.008	-2.03	0.04
Interaction terms	Goal *Subject	0.76	0.49	1.55	0.12
	Goal*Gender	-0.23	0.48	-0.47	0.64
	Subject*Gender	-0.006	0.48	-0.01	0.99
	Subject*Gender*Goal	0.32	0.97	0.33	0.74
Control	Use of connective word	0.88	0.30	2.93	0.004
	Likelihood	---	---	---	---
Random Effects	Participant	*			
	Participant by Subject vs. non-Subject	*			
	Participant by Goal vs. Source	*			

Note. T-values for predictor variables and interaction terms indicate their significance. Control variables that were significant in the main effects model were included, and T-values indicate their significance in this model. Dashed lines for control variables indicate the variable was not significant in the main effects model and thus was not included here. Random effects are noted with asterisks if included.

### *Connective use effects*

Participants' use of connective words was also analyzed. Use of a connective was hypothesized to be an indicator of the participants' use of the discourse context. If participants conceptualized the two events as a related unit, this strategy might allow them to use the description of the first event to lead into the second event. Conceptualizing of the two events as a whole would encourage them to think about the relationships between the two, potentially increasing their use of words like (and then, next, after that) to emphasize their linkage.

Participants in Experiment 4 used far more connectives (194) than participants in the other experiments with comparable numbers of participants (11 in Experiment 1, 9 in Experiment 2). This was likely due to both the verbal modality of response and possibly their use of the discourse context.

The main effects analysis of participants' use of connectives (*after, afterwards, and, next, now, then.*) found that participants use more connectives when talking about Sources,  $F(1,377)=4.72, p=.03$ , and the Subject effect was estimated in the hypothesized direction,  $F(1,19)=2.78, p=0.11$  (see Table 9). These effects were not qualified by any interactions (see Table 10 for interaction model details).

Table 9

Experiment 4 predictor, control variables and random effect in the main effects model for connective use

Critical Predictors	Variable	Estimate	Error	t-value	p-value
	Goal vs. Source	-0.67	0.31	-2.17	0.03
	Subject vs. non-Subject	0.56	0.34	1.67	0.11
	Same gender vs. Different gender	-0.29	0.29	-1.00	0.32
	Order	-0.012	0.009	-1.27	0.20
Control Variables	List	---	---	---	---
	Mention Other Person	---	---	---	---
	Referent on right in picture	0.77	0.30	3.54	0.012
	Relatedness	-0.33	0.23	-1.43	0.15
	Likelihood	---	---	---	---
	Count of words	---	---	---	---
	Ease	---	---	---	---
	Referential form (pro/zero vs name)	0.73	0.34	2.12	0.03
	Disfluency	-1.26	0.45	-2.79	0.005
	Latency to begin speaking	---	---	---	---
Random Effects	Participant	*			
	Participant by Subject vs. non-Subject	*			
	Participant by Goal vs. Source	Not positive definite			

Note. T-values for predictor variables indicate their significance. Control variables with *t*-values >1.5 in the control model were included in the main effects model and their values in the main effects model are given here. Dashed lines for control variables indicate the variable was not significant in the control model and thus was not included. Random effects are noted with asterisks if included.

Table 10

Experiment 4 predictor variables, control variables, interaction terms, and random effects in the interaction model for connective use

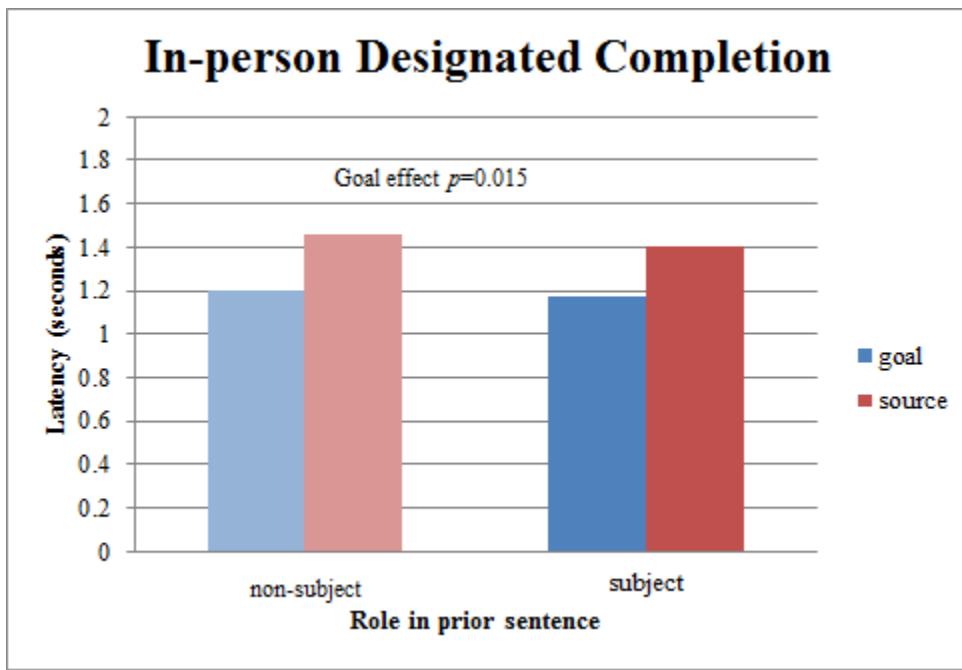
	<b>Variable</b>	<b>Estimate</b>	<b>Error</b>	<b>t-value</b>	<b>p-value</b>
<b>Critical Predictors</b>	Goal vs. Source	-0.77	0.30	-2.58	0.01
	Subject vs. non-Subject	0.56	0.34	1.66	0.11
	Same gender vs. Different gender	-0.42	0.28	-1.50	0.13
	Order	-0.01	0.009	-1.44	0.15
<b>Interaction terms</b>	Goal *Subject	0.31	0.56	0.56	0.57
	Goal*Gender	0.49	0.56	0.89	0.37
	Subject*Gender	-0.10	0.56	-0.19	0.85
	Subject*Gender*Goal	0.62	1.11	0.55	0.58
<b>Control Variables</b>	Referent on right in picture	0.67	0.30	2.21	0.03
	Relatedness	---	---	---	---
	Disfluency	-1.22	0.46	-2.65	0.008
	Referential form (pro/zero vs name)	0.66	0.35	1.92	0.05
<b>Random Effects</b>	Participant	*			
	Participant by Subject vs. non-Subject	*			
	Participant by Goal vs. Source		Not positive definite		

*Note.* T-values for predictor variables and interaction terms indicate their significance. Control variables that were significant in the main effects model were included, and T-values indicate their significance in this model. Dashed lines for control variables indicate the variable was not significant in the main effects model and thus was not included here. Random effects are noted with asterisks if included.

### *Latency effects*

If Goal continuations are easier to plan and produce, we would expect to see an effect on participants' latencies to begin speaking. The latency to begin speaking was calculated as the time it took the participant to begin speaking after the detective had finished the description of the first picture.

Indeed, the latencies to begin speaking after the detective had finished were shorter when the participant was referring to a Goal as opposed to a Source,  $F(1,362)= 5.97$ ,  $p=0.015$ . There was no such effect of referring to the Subject versus the non-Subject,  $F(1,19)=0.08$ ,  $p=.78$ , and no interaction between the two ( $p=0.97$ ). Latency data can be seen in Table 11 and Figure 16. These effects were not qualified by any interaction, although there was an effect approaching significance in the 3-way interaction between gender, Goal-hood, and Subject-hood, as can be seen in Table 12. As the main effect of Goal and the lack of Subject effect both remained consistent in the presence of this interaction it was not investigated further.



*Figure 16.* Latency to begin speaking by grammatical and semantic roles in prior sentence in Experiment 4

Table 11

Experiment 4 predictor and control variables in the main effects model for latency to begin speaking

	<b>Variable</b>	<b>Estimate</b>	<b>Error</b>	<b>t-value</b>	<b>p-value</b>
<b>Critical Predictors</b>	Goal vs. Source	-0.07	0.027	-2.44	0.015
	Subject vs. non-Subject	-0.005	0.02	-0.28	0.78
	Same gender vs. Different gender	0.02	0.02	1.10	0.27
	Order	0.001	0.0006	1.84	0.07
<b>Control Variables</b>	List	---	---	---	---
	Mention Other Person	---	---	---	---
	Use of connective word	---	---	---	---
	Referent on right in picture	0.07	0.02	3.52	0.0005
	Relatedness	-0.05	0.01	-3.24	0.001
	Count of words	---	---	---	---
	Likelihood	0.005	0.05	0.11	0.91
	Ease of verb retrieval	-0.07	0.03	-2.18	0.03
	Disfluency	0.21	0.03	6.96	<.0001
	Referent form choice (pro/zero vs name)	-0.004	0.02	-0.20	0.84
<b>Random Effects</b>	Participant	*			
	Participant by Subject vs. non-Subject	*			
	Participant by Goal vs. Source	Not positive definite			

*Note.* T-values for predictor variables indicate their significance. Control variables with *t*-values >1.5 in the control model were included in the main effects model and their values in the main effects model are given here. Dashed lines for control variables indicate the variable was not significant in the control model and thus was not included. Random effects are noted with asterisks if included.

Table 12

Experiment 4 predictor variables, control variables, interaction terms, and random effects in the interaction model for latency to begin speaking

	<b>Variable</b>	<b>Estimate</b>	<b>Error</b>	<b>t-value</b>	<b>p-value</b>
<b>Critical Predictors</b>	Goal vs. Source	-0.06	0.02	-3.27	0.0012
	Subject vs. non-Subject	-0.007	0.02	-0.40	0.70
	Same gender vs. Different gender	0.02	0.02	1.09	0.28
	Order	0.0012	0.0006	1.98	0.05
<b>Interaction terms</b>	Goal *Subject	0.001	0.035	0.03	0.97
	Goal*Gender	0.04	0.037	1.17	0.24
	Subject*Gender	-0.007	0.036	-0.19	0.85
	Subject*Gender*Goal	0.12	0.07	1.66	0.10
<b>Control Variables</b>	Referent on right in picture	0.07	0.2	3.56	0.0004
	Relatedness	-0.05	0.015	-3.02	0.002
	Disfluency	0.21	0.03	6.92	<.0001
	Ease of verb retrieval	-0.08	0.03	-2.33	0.02
	Referent form choice (pro/zero vs name)	---	---	---	---
	Likelihood	---	---	---	---
<b>Random Effects</b>	Participant	*			
	Participant by Subject vs. non-Subject	*			
	Participant by Goal vs. Source		Not positive definite		

*Note.* T-values for predictor variables and interaction terms indicate their significance. Control variables that were significant in the main effects model were included, and T-values indicate their significance in this model. Dashed lines for control variables indicate the variable was not significant in the main effects model and thus was not included here. Random effects are noted with asterisks if included.

Table 13

## Correlations between predictors

Ease of verb	Ease of verb	Likelihood $r(422)=-0.26, p=.60$	Relatedness Z $r(422)=-.085, p=.081$ $r(422)=.198, p<.0001$
Likelihood			

*Post-experiment questionnaire*

A post-experiment questionnaire (see Appendix H) was used to achieve a couple of goals. First, we were interested in determining participants' prior experience with the Clue game. We hypothesized that participants' familiarity with the game would allow them to create richer mental models and draw on their experiences with the game. This use of prior experience and enriched mental models would perhaps allow participants to utilize the discourse more.

We were also interested in motivating participants' engagement by encouraging them to try to figure out who had solved the crime. Participants were told at the beginning of the experiment that at the end they would try to help the detective figure out who the culprit was. The post-experiment questionnaire allowed us to gauge whether participants had paid attention and attributed the action and a motive to someone.

All 20 participants correctly identified Sir Barnes as having been the character that was killed. All of the participants identified the motive as having something to do with Sir Barnes' affair with the maid. Eighteen identified the butler as having been the killer, either alone (16) or with the chef (2). One participant thought Lady Mannerly killed Sir Barnes, and one thought it was the maid. Sixteen participants thought he had been killed with the rifle, one thought it was a combination of the rifle and a drink, one thought it was the rifle or a knife, one thought it was poison, and one thought it was by the knife. Eighteen of the 20 participants said they were familiar with the game Clue, and seventeen reported having played it. Eight participants reported

having seen the movie. Participants reported varied possibilities for what the experiment had been about, but none identified the specific linguistic manipulations. Given that all participants correctly identified the victim of the crime and the majority (18) correctly identified the culprit, participant engagement during the task was deemed to be satisfactory.

### Discussion

As found in Experiments 1 and 2, participants in Experiment 4 used more pronouns or zeros when talking about Subjects of the prior sentence as opposed to non-Subjects, and Goals of the prior sentence as opposed to Sources. These findings were the same as those found with the less interactive, standard sentence continuation design and suggest that this effect can be found in a variety of experimental settings, and was not related to a particular feature of the prior design, such as writing responses or unlimited planning time.

One interesting finding was that participants used more connectives (*then, and, after, etc.*) when talking about Sources of the prior sentence as opposed to Goals, and Subjects of the prior sentence as opposed to non-Subjects. This may seem odd at first, given that participants were also always describing the event that happened next, so a focus on the end state of the event, and thus the Goal, would be expected. This might lead to the expectation that participants would use connectives more when referring to the Goals, to emphasize the next-event continuation by the most likely referent. However, perhaps participants were using these constructions precisely because the Source continuation was less expected. That is, perhaps the use of a connective helped establish a relationship between two events that flowed less naturally. Participants used far more connectives in this experiment than in the online experiments, probably due primarily to the verbal nature of their response, but also perhaps reflecting a greater use of the discourse context.

This experiment also addressed the question of what mechanism might underlie the bias to use reduced forms for goals by examining the latency of utterance onset. If the same forces that caused participants to begin speaking more quickly also caused them to use more pronouns or zeros, we would expect to see a relationship between these two variables. If, instead, different factors influenced how quickly participants began speaking and what form of referring expression they chose, we would not see that utterances that began sooner also made use of reduced forms at a higher rate. These two possibilities are related to the mechanisms under consideration. Under the facilitation mechanism, predictability information doesn't directly influence accessibility but speeds planning and production processes. That mechanism would be supported by evidence that measures of planning affected reference form choice.

The latency data from this experiment suggested that while participants were faster to begin speaking when talking about the more predictable referent (the Goal), this speed of production was not related to choice of referring expression. However, latency to begin speaking was not a significant predictor of pronoun/zero choice, as indicated in the main effects model for referential form choice. Nor were other measures of planning difficulty: whether or not participants were disfluent was not related to referential form, nor was the measure of verb retrieval ease. This pattern of results provides some qualified evidence against the facilitation mechanism, by which predictability influences referential form through production planning.

This evidence is not definitive, though, because the latency measure was a somewhat rough measure of participants' planning. The latency was measured from the end of the detective's speech until the beginning of the participant's speech. However, participants had, at that point, been examining the pair of pictures and had heard the description of the first event.

Their planning, therefore, had been going on for several seconds, and they had already been able to examine the pictures and establish the relationship between them.

The measure of latency used in this study was able to detect differences between the Goal and Source items as groups, but was perhaps not sensitive enough to detect the more minor differences, if they existed, within these classes between different referential forms. A more sensitive latency measure would perhaps measure from the onset of the second picture until the beginning of the participant's description. However, as one aim of this study was to have participants conceptualize the pictures as a group, that measure was not possible given the other goals of this study.

It is also possible that latency to begin speaking did not have a relationship with form of referring expression because of the other speech planning considerations. Participants must balance their speech onset with utterance planning considerations. When participants begin an utterance with a pronoun they must have more of it planned out, as the short length of the pronoun will not allow them to be incrementally planning the following words. This pressure may force them to begin speaking slightly later than they possibly could have, although perhaps the pronoun was retrieved and ready to be produced. In contrast, when beginning an utterance with a name, participants can still be planning the next words while speaking. Thus, it is possible that pronouns were retrieved and ready to be produced sooner than names, but the different planning constraints associated with both choices resulted in roughly similar latencies to begin speaking.

## EXPERIMENT 5: EVENT PREDICTABILITY

### *Motivation*

Experiments 1-4 established the effect of thematic roles on referring expressions within both the traditional sentence continuation paradigm and in a new, interactive event-retelling task. Experiment 4 provided qualified evidence that predictability was not influencing referential form via a facilitation mechanism.

Experiment 5 was conducted to clarify that predictability was driving the effects seen. The next-mention bias in the free-completion experiment was taken as a proxy for predictability, but we wanted to confirm these predictability judgments. We also wanted to probe the specific kind of predictability participants were sensitive to. In Experiment 5 we evaluated next-event predictability: whether participants had predictability expectations about certain events. The characters who featured in these events (Goals or Sources) likely were an important part of participants' predictability calculations, but so too may be the events themselves.

Participants were provided with pairs of sentences (see Appendix H) and depictions of the events from the storyline. They were asked to rate (1) how related the second event depicted/described in each pair was to the first, and (2) how predictable the second event depicted/described in each pair was to the first. The questions are slightly different from each other, because the predictability question is essentially asking participants to gauge the event against all possible events that might have followed the first. The relatedness question may be less demanding to answer: it simply requires participants to consider if this event makes sense, given the first, rather than weighing it against some hypothetical predictions.

## Method

### *Participants*

21 undergraduates completed the task, all for course credit. One participant was excluded for not being a native English speaker, leaving 20 participants who were included in the analysis. In order to be included participants needed to be native speakers of English, have normal or corrected-to-normal vision, and not have participated in a similar experiment in the lab.

### *Materials and Design*

Participants viewed the set of 53 pairs of sentences and pictures described above, which included 24 critical items. Participants were recruited through SONA and completed the task for class credit. The sentences and pictures were presented to participants with a computerized survey through Qualtrics. There were two versions of the survey, which differed only in the first sentences of the critical items. Participants therefore saw each item in one of the two conditions, but both conditions across different items. The sentence stimuli for the rating study can be found in the Appendix. Both versions of an item from this rating study are given in Figures 17 and 18.

Sir Barnes received a painting of the two of them from Lady Mannerly. Sir Barnes threw the painting in a closet.

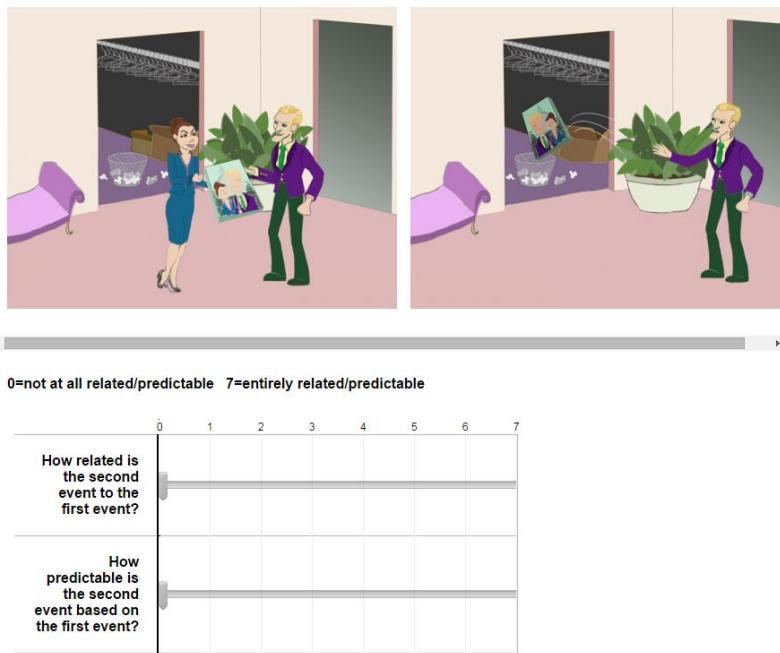


Figure 17. Subject-Goal continuation version of an item from Experiment 5

Lady Mannerly gave a painting of the two of them to Sir Barnes. Sir Barnes threw the painting in a closet.

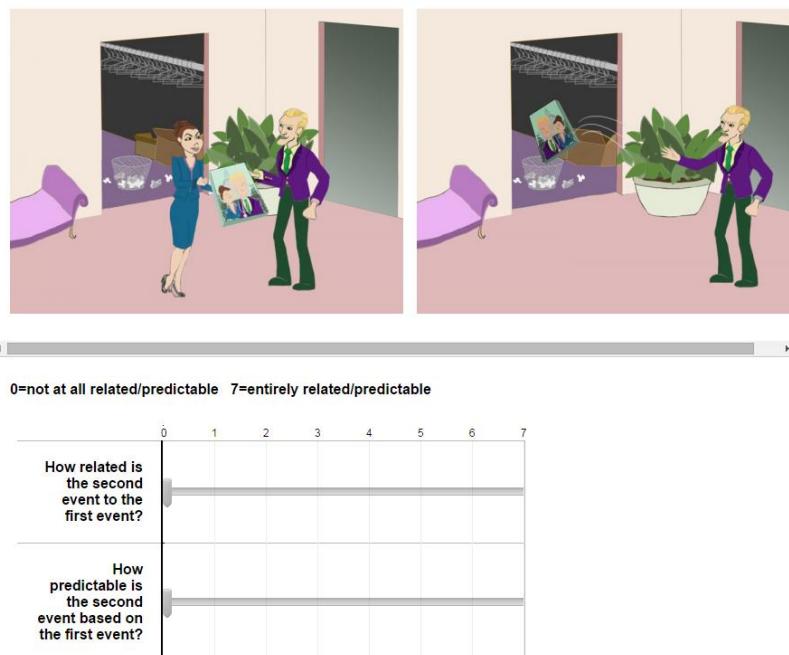


Figure 18. non-Subject Goal continuation version of an item from Experiment 5

### *Procedure*

After consenting to the study, participants viewed a narrated slideshow that introduced them to the storyline and the characters. The transcript of the slideshow is given in Appendix F. The slideshow told them that they were a tabloid photographer and that they had been taking pictures at the home of a wealthy British couple. While they were at the house a murder occurred, and they had not yet had a chance to review the photographs they had taken. This was the same slideshow as was shown to participants in Experiment 4 as background, but critically, in Experiment 5 participants did not then preview the pictures.

After viewing the background slideshow participants were told that they would now be seeing the pictures they had taken, for the first time, in order. Participants were then presented with each pair of pictures and the sentences that described them, via a computerized survey. Participants were instructed to rate the second event depicted and described on (1) how related it was to the first event and (2) how predictable it was, based on the first event. Participants used a scale to provide responses to the survey that were recorded as whole numbers from 0-7. A lab confederate watched the participants complete the task to ensure quality of data.

### *Analysis*

Two values were collected for each item and every participant, relatedness and predictability. The data on both of these scales for each participant were transformed into z-scores, using that participant's mean and standard deviation for that rating (relatedness or predictability). This transformation was done to ensure uniformity of scale, as participants may have utilized the range of responses differently.

The relatedness ratings and predictability ratings were analyzed within separate models. The same 3-step model process (control model, main effects model, interaction model) as in

earlier experiments was used for both variables. SAS proc mixed was used, as the outcome measures (z-score of rating) were continuous. No random effects were included as the outcome measure was a Z-score.

After testing the control variables and retaining those with  $t$ -values greater than 1.5, the final models were built and contained any significant control variables as well as the critical predictors of semantic predictability (Goal or Source-continuations) and referential predictability (Subject or non-Subject-continuations). Gender was not included in these models as a predictor variable as it was not predicted to affect the ratings of event relatedness and was not of theoretical interest. Any anecdotal features of the items, such as gender make-up, will be reflected in participants' ratings.

## Results

### *Semantic and grammatical role effects on ratings*

Participants rated the Goal continuations as being more related  $F(1,475)=20.22, p<.0001$  than the Source Continuations, as shown in Figure 19. Interestingly, participants did not judge continuations of the Subject of the prior phrase to be more related  $F(1,475)=0.04, p=.84$  than continuations of the non-Subject (see Table 14). This effect was not qualified by an interaction between thematic and grammatical roles (see Table 15).

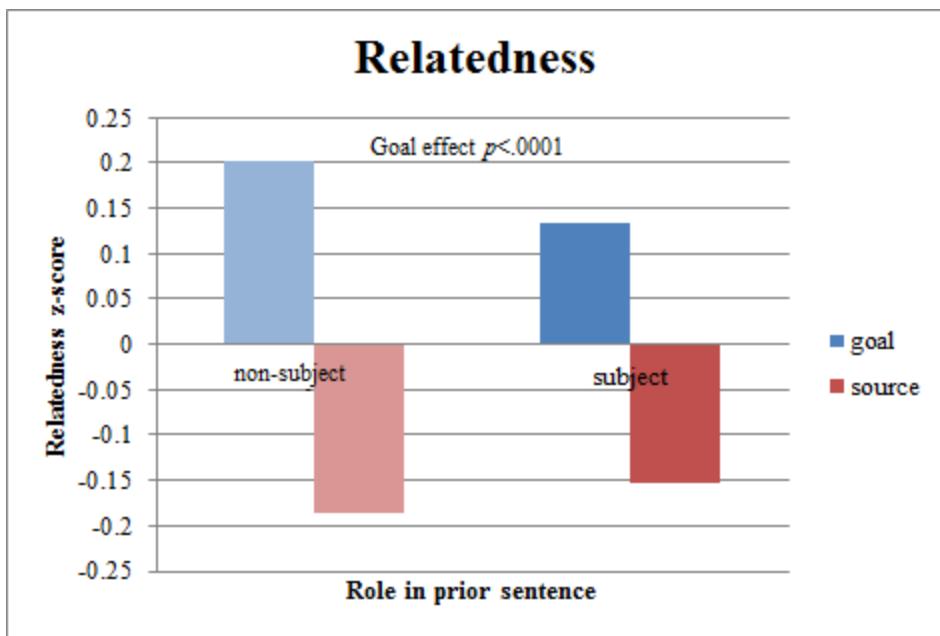


Figure 19. Z-score ratings of relatedness by role in prior sentence in Experiment 5

Table 14

Experiment 5 predictor and control variables in the main effects model for Relatedness

Critical Predictors	Variable	Estimate	Error	t-value	p-value
	Goal vs. Source	0.41	0.09	4.50	<.0001
	Subject vs. non-Subject	-0.02	0.09	-0.20	0.84
	Order	0.008	0.003	2.55	0.01
Control	List	---	---	---	---
	Referent on right in picture	0.008	0.003	2.55	0.01

Note. Control variables with t-values  $> 1.5$  in the control model were included in the main effects model and their values in the main effects model are given here. Dashed lines indicate the variable was not significant in the control model.

Table 15

Experiment 5 predictor variables, control variables, interaction terms, and random effects in the interaction model for Relatedness

Critical Predictors	Variable	Estimate	Error	t-value	p-value
	Goal vs. Source	0.41	0.09	4.49	<.0001
	Subject vs. non-Subject	-0.02	0.09	-0.20	0.84
	Order	0.008	0.003	2.55	0.01
Interaction	Goal *Subject	-0.10	0.17	-0.59	0.56
Control	Referent on right in picture	0.34	0.09	3.67	0.0003

Note. T-values for predictor variables and interaction terms indicate their significance. Control variables that were significant in the main effects model were included, and T-values indicate their significance in this model.

The Goal effect was in the predicted direction but was not significant,  $F(1,475)=3.70$ ,  $p=.055$  as shown in Figure 20 (also see Table 16). As with the relatedness judgments, participants did not find continuations of the Subject of the prior phrase to be more predictable  $F(1,475)=0.45$ ,  $p=.50$  than continuations of the non-Subject. These main effects were not qualified by any interactions (see Table 17).

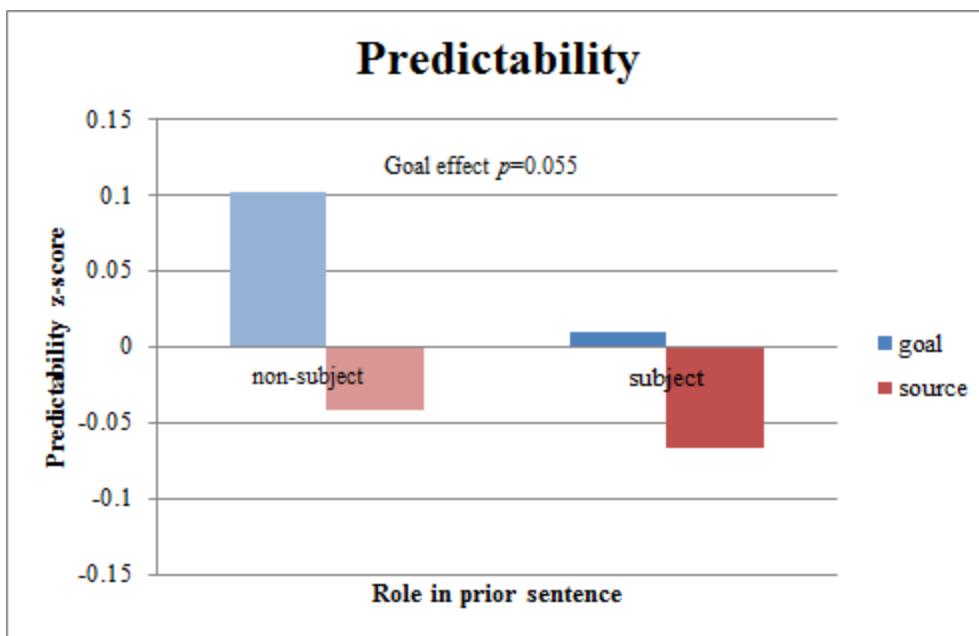


Figure 20. Z-score ratings of predictability by role in prior sentence in Experiment 5  
Table 16

Table 16

Experiment 5 predictor and control variables in the main effects model for Predictability

Critical Predictors	Variable	Estimate	Error	t-value	p-value
Critical Predictors	Goal vs. Source	0.18	0.09	1.92	0.055
	Subject vs. non-Subject	-0.06	0.09	-0.67	0.50
	Order	0.007	0.003	2.31	0.02
Control	List	---	---	---	---
	Referent on right in picture	0.41	0.09	4.47	<.0001

*Note.* Control variables with t-values >1.5 in the control model were included in the main effects model and their values in the main effects model are given here. Dashed lines indicate the variable was not significant in the control model. T-values for predictor variables indicate their size in the main effects model as well

Table 17

Experiment 5 predictor variables, control variables, interaction terms, and random effects in the interaction model for Relatedness

Critical Predictors	Variable	Estimate	Error	t-value	p-value
Critical Predictors	Goal vs. Source	0.18	0.09	1.92	0.06
	Subject vs. non-Subject	-0.06	0.09	-0.67	0.50
	Order	0.007	0.003	2.31	0.02
Interaction	Goal *Subject	-0.07	0.17	-0.38	0.70
Control	Referent on right in picture	0.41	0.09	4.47	<.0001

*Note.* T-values for predictor variables and interaction terms indicate their significance. Control variables that were significant in the main effects model were included, and T-values indicate their significance in this model.

## Discussion

Experiment 5 revealed that participants found Goal continuations to be more related and predictable than Source continuations. They did not find events featuring prior-Subjects to be more related or predictable than events featuring prior non-Subjects.

The results of this rating study confirmed and extended the results of Experiment 3. In addition to preferring to refer to Goals, participants found Goal-continuation events to be more related and predictable than Source-continuation events. We did not see, however, that participants found Subject continuations to be more predictable than non-Subject continuations. Topicality effects are driving the differences in pronoun/zero rates between Subject and non-Subject continuations, while predictability effects are influencing Goal and Source referential forms.

One important point to consider is what the relationship between these predictability ratings and the actual use of predictability during speech is. Presumably the participants in this task were making judgments about the relationships between the two events, while predictability during normal speech likely has its effects over a much shorter timeframe and on less obvious choices. These data tell us something important about how participants rate these continuations when they are given the time to consider them, but these predictability ratings may not map directly onto the predictability judgments informing choices during online production and comprehension. However, the pattern of these results (Goals preferred over Sources, Subjects not preferred over non-Subjects) is the same as the other measure of predictability (next-mention) that was a more accurate real-time measure of predictability information.

## EXPERIMENT 6: REFERENT PREDICTABILITY

### Method

#### *Motivation*

Experiment 6 was conducted as a specific test of referent predictability, that is, how predictable the character itself was to be continued. By contrast, Experiment 5 examined event predictability: participants were rating the events for how likely they were to occur. This is an important distinction to make, because it is possible that participants' judgments about these two predictabilities could be different. The results of Experiment 5 may be attributable to predictability judgments made about the characters, but there is also the possibility that there were other factors about the events depicted, in addition to the person performing the action, that influenced participants' likelihood ratings. Indeed, given the unexpected effects of gender and the interaction between thematic role and gender, it seems likely that participants' responses were strongly influenced by the actual events.

In addition, all the pictures in Experiment 5 were of the next event. This necessarily acts against the Source bias for explanation or background continuations, and is in line with the Goal bias for next-event mentions. Given that the events shown in Experiment 5 all depicted and described the next mention, participants may have been unduly biased to find them to be more predictable for Goals as opposed to Sources.

If, instead, some of the pictures featuring Source continuations had depicted the precipitating events, or the characters' motivations, participants may have rated Source items as more related than they did. We were interested in whether the results of Experiment 5 were artifacts of the next-event feature of the pictures, due to anecdotal features of the pictures

themselves, or were actually driven by participants' intuitions about which referent was likely to be mentioned next (regardless of the content of that utterance).

Experiment 6 was done in person. Participants were provided with the first picture and sentence of each pair from the storyline. They were then shown individual pictures of the two characters in that item and asked to choose which was most likely to be talked about next.

### *Participants*

20 undergraduates completed the task, all for course credit. In order to be included participants needed to be native speakers of English, have normal or corrected-to-normal vision, and not have participated in a similar experiment in the lab.

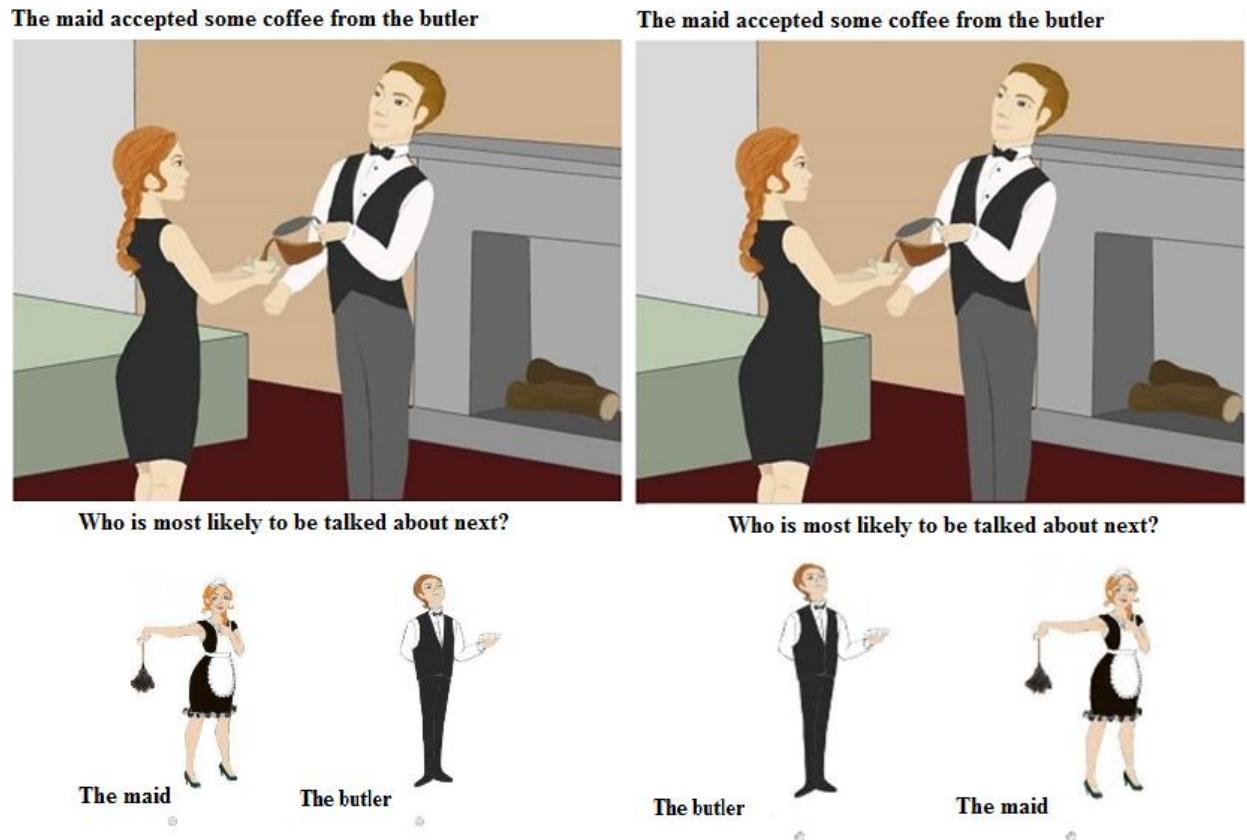
### *Materials and Design*

Participants viewed the first sentence and picture of each of the 53 pairs of sentences and pictures described above. Participants were recruited through SONA and completed the task for class credit. The sentences and pictures were presented to participants with a computerized survey through Qualtrics.

Four versions of the survey were created, two from each of the previously described lists. For the critical items, the position of the two characters in the response option was balanced across the two versions of each list, such that a character appeared on the right side (for the response) in one list for a particular item, and appeared on the left side (for the response) in the other list for that particular item. Half of the time the Subjects was presented on the same side in the response options (R or L) as it was in the picture, and half of the time they were on different sides.

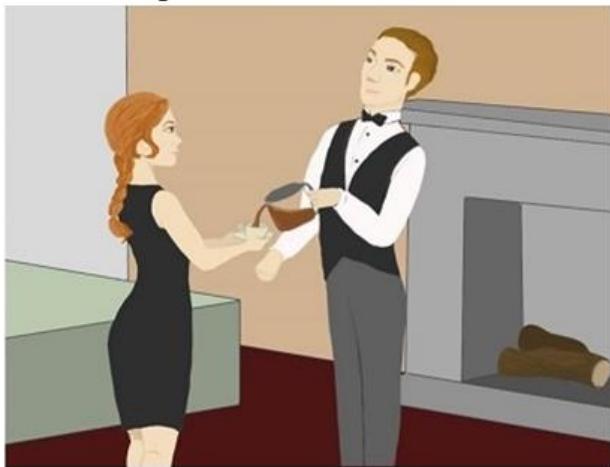
The filler items were similarly counterbalanced, such that the character given on the right in the response in one version of each list was given as the left response option in the other

version. Only eight of the filler followed the Subject and non-Subject framework. An additional eleven of the fillers had compound Subjects, so for these items one of the characters was put on the left in the response and one on the right. For the ten fillers in which only one person was mentioned, a second character was substituted as the second option for who might be mentioned next. This character was selected to be someone who was not presented in the next item. All four versions of an item in this study are given in Figures 21 and 22.



*Figure 21. List A Versions 1 and 2 of an item from Experiment 6*

The butler brought some coffee to the maid



Who is most likely to be talked about next?

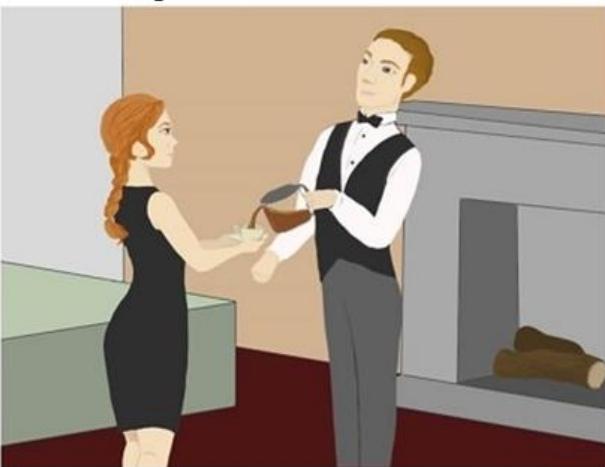


The maid



The butler

The butler brought some coffee to the maid



Who is most likely to be talked about next?



The butler



The maid

Figure 22. List B Versions 1 and 2 of an item from Experiment 6

#### Procedure

The procedure was the same as for Experiment 5. After consenting to the study participants viewed the narrated slideshow, which told them they were a tabloid photographer and had been secretly taking photos of a wealthy British couple in their home. As in Experiment 5 they did not preview their pictures.

After viewing the background slideshow, participants were told that they would now be seeing the pictures they had taken, for the first time, in order. Participants were then presented, with one picture at a time. The description of the picture was given at the top and below they were asked to choose which character they thought would be most likely to be talked about next. After selecting that character they moved on to the next picture. A lab confederate watched the participants complete the task to ensure quality of data.

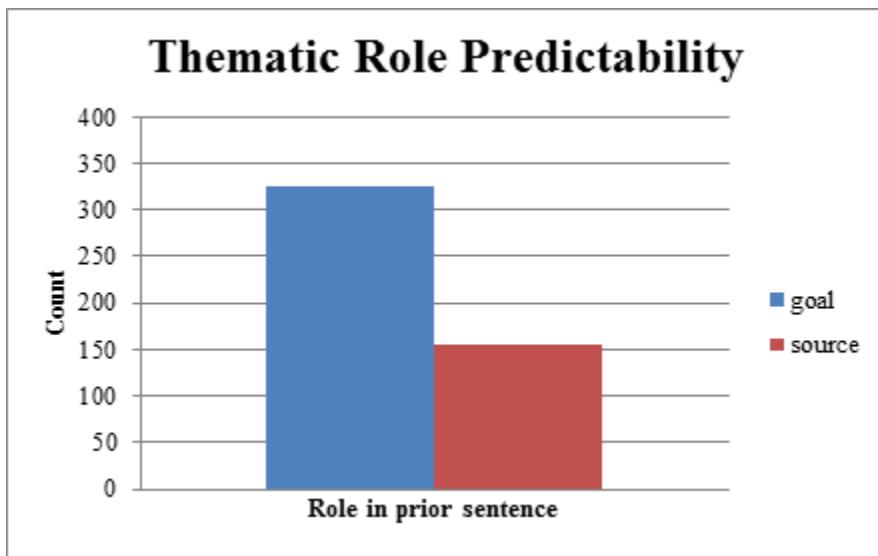
## *Analysis*

One value was collected for each item and every participant: which character they thought was most likely to be talked about next. This data was analyzed with SAS proc glimmix. In order to determine who was most likely to be talked about, two random intercept models were run with random intercepts for participants. This allowed us to determine whether participants had a significant preference to select either the Goal or the Source as the more likely character to be continued. The same procedure was done with Subject or non-Subject chosen as the outcome measure and a random intercept for person, to determine whether participants had a preference for either Subject or non-Subject to be continued.

## Results

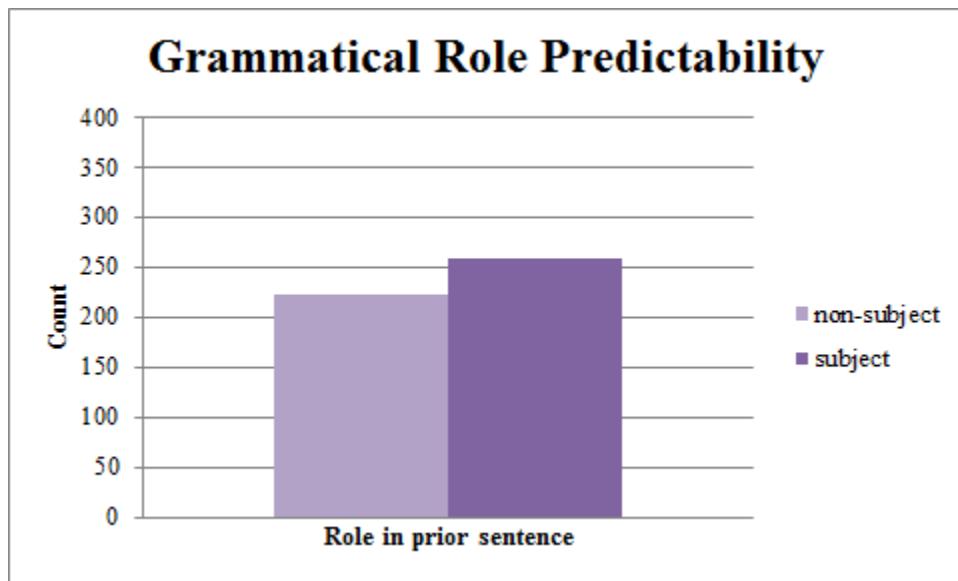
### *Semantic and grammatical role effects on ratings*

Participants rated Goals as more likely to be referred to next than Sources. The probability that the average participant would choose the goal as being the next person referred to was 71%. The intercept of the random intercept model was significant,  $t(19)=3.71, p=.0015$ . The probability of choosing the Goal was calculated using the intercept estimate of 0.8750:  $p=\text{exponential function}(0.8750)/(1+(\text{exponential function}(0.8750)))= 0.706$ . The counts of Goal and Source referents chosen can be seen in Figure 23.



*Figure 23.* Count of likelihood judgments for Experiment 6 by semantic role in the prior sentence

Participants did not rate Subjects as more likely to be referred to next compared to non-Subjects. The probability of choosing the Subject, for an average person, was 54%. The intercept of the random-intercept only model was not significant  $t(19)=1.45, p=.16$ , and the probability was calculated using the intercept estimate 0.1507,  $p= \text{exponential function}(0.1507)/(1+\text{exponential function}(0.1507))$ . The counts of Subjects and non-Subjects chosen as the most likely next mentioned character are given in Figure 24.



*Figure 24.* Count of likelihood judgments for Experiment 6 by grammatical role in the prior sentence

### Discussion

Experiment 6 found an effect of referent predictability. Participants found referents in the Goal role to be more likely to be continued than referents in the Source role. These results are consistent with those of Experiment 5, which examined event predictability. While numerically it appears that participants may have a small preference for continuing to talk about Subjects, the probability of a participant choosing the Subject as compared to the non-Subject was only slightly above 50% and was not statistically significant.

These results confirmed, with an explicit next-mention judgment, participants' biases for continuations to be about Goals as opposed to Sources. The results that people find Goal referents to be more predictable are in line with the findings of Experiments 3 and 5: participants also prefer to refer to Goals, and they rate events involving Goals as more predictable. The Goal bias was even stronger in Experiment 3 than Experiment 6: participants chose to refer to Goals 82% of the time in Experiment 3. By comparison, in Experiment 6 they rated Goals as the more

likely continuation 71% of the time. This indicates that implicit biases and explicit ratings are similar, but implicit biases may be even stronger than explicit judgments.

The Subject-continuation findings from Experiment 6 are also consistent with those of Experiments 3 and 5. Although numerically there may appear to be a small preference for Subject as opposed to non-Subject continuations, participants chose the Subject as the most likely continuation only 54% of the time. In Experiment 3 participants chose to refer to the Subject 48% of the time, and in Experiment 5 they did not rate Subject continuations to be more predictable or related than non-Subject continuations.

## CHAPTER 5: GENERAL DISCUSSION

The major finding in this project was that semantic role biases significantly affected pronoun/zero rates. This effect was found in all three of the experiments in which we were able to collect a substantial amount of continuations in both the preferred (Goal continuation) and non-preferred (Source continuation) conditions. The only experiment in which we did not see this effect was Experiment 3. Here, participants did show a trend toward greater pronoun/zero use for Goal referents than Source referents. However, this effect was not statistically reliable due to the fact that they dispreferred mentioning Source referents overall. A summary of the effects of the critical predictors and control variables from Experiments 1-4 is given in Table 18.

Table 18

Summary of the critical and control predictor effects on referential form from Experiments 1-4

		Direction	Exp. 1 Designated	Exp. 2 Renamed	Exp. 3 Free	Exp. 4 In person
Critical Predictors	Goal	go>src	X	X		X
	Subject	subj>non-subj	X	X	X	X
	Gender	diff>same	X	X	X	X
Control variables	Order	later>earlier	X	X	X	earlier>later
	Mention other character	Mentioned>Not mentioned	X			
	Connective	Connective>No connective				X

The results of this set of studies differ from those previously reported, which found no effects of semantic role biases on pronoun/zero rates (with the exception of Arnold, 2001), but robust effects of topicality. These data are inconsistent with the narrow definition of topicality, but are consistent with a broader conceptualization in which predictability is a component of topicality. One contributing factor to the presence of the effect in the current experiments is the use of Goal-Source verbs, which were hypothesized to be more likely to show an effect than Implicit Causality verbs. The thematic biases of Goal-Source verbs are related to the expected behavior of participants after something has been transferred. This transfer event should be relatively simple to incorporate into a mental model and make predictions about. These predictions do not depend on any consideration of psychological states or motivations of the referents, but simply the usual outcome after an object changes hands.

Indeed, it has been hypothesized (Tatone, Geraci, & Csibra, 2015) that humans have a dedicated action schema for the action of giving, which allows the efficient representation of such events. This action schema is argued to exist because giving actions are such central components of human interactions, and played a fundamental role in evolutionary history. The existence of such a schema would certainly facilitate the representation of transfer verbs, and would perhaps require fewer working memory resources than representing a verb that described an emotional state.

By comparison, Implicit Causality verbs require modeling participants' mental states. This leaves multiple possibilities for difficulty in creating mental models and predictions, as participants must hypothesize about the characters' emotions. Given the resources necessary to create the mental model of the event, participants may have fewer resources available to represent the referents themselves. Predictability information also may not be available to direct

attention to one referent over another, or may not be available in time for form selection, given the steps involved in determining the character's mental states. This would result in no difference in the rates of reduced forms used for the more predictable referent versus the less predictable referent.

This comparison also suggests another reason that participants prefer to refer to the Goal referents, in addition the possible explanation of the focus on the end state of the event. If participants are forced to talk about the Source they don't have the option of talking about what happened next with the object that was transferred. They must instead consider the motivations of the Source referent, and discuss what precipitated the transfer event. This sort of calculation may be more mentally taxing than simply describing what a referent did next with a particular object. If so, this might contribute to the finding that people use fewer reduced forms for Sources.

Whether the semantic biases of Implicit Causality verbs affect referential form choice within this story-continuation paradigm or not is the subject of further study. If they do not, the interesting question of what characteristic of Goal-Source verbs contributes to predictability emerges. Implicit Causality and Goal-Source verbs differ on many dimensions, any (or all) of which might be critical for the effect of predictability.

A final difference between this set of studies and previous was the exclusion criteria related to referential form variation. Previous studies included participants regardless of linguistic form variation, and the current study only included participants who used some variation in their expressions. As discussed earlier, our results mirrored previous findings, which found an effect for grammatical role but not semantic role, when the sample of the first 20 participants was analyzed without regard to linguistic variation. In order to find the effects of

semantic role predictability, which are relatively small as compared to grammatical role effects, it is necessary to only analyze those participants who do not adopt a general approach to the task (i.e., refer to everybody with pronouns/zeros, refer to everybody with proper names) but show sensitivity to the manipulations.

Although participants used more reduced forms for Goals versus Sources, there was not a straightforward relationship between predictability and reference form in general. Predictability did not account for the increased use of reduced forms to refer to Subjects versus non-Subjects. In the in-person study the difference in the rate of pronoun/zero use between Subject (50%) and non-Subjects (24%) was roughly similar to the rates between Goals (45%) and Sources (25%). However, the predictability ratings did not show this same pattern. Participants chose to talk about Goals as opposed to Sources (83% of the time in the free completion study), and no such effect was found for Subjects versus non-Subjects (percentage of choosing to talk about the Subject were 48%). Participants rated Goals as being more likely to be referred to next as compared to Sources, (71%), but not Subjects compared to non-Subjects (54%). Events that were Goal-continuations were found to be more related ( $p=.0001$ ) than Source-continuations, but this was not the case for those that were Subject-continuations ( $p=0.8$ ) as opposed to non-Subject-continuations.

These results are interesting because it has been claimed that speakers continue talking about referents that are in Subject position (Brennan, 1995), and indeed some evidence has been found to support this (Arnold, 1998 dissertation). This pattern of continued mention should make Subject referents more likely to be referred to next, and thus more predictable. One caveat is that these continuations are likely modulated by the coherence relations between phrases, a topic which will be discussed in greater detail later. However, if Subjects are generally repeatedly

mentioned, it is puzzling that the participants in this study did not show this bias: they did not rate Subjects as more likely to be referred to when compared to non-Subjects. A visual examination of the likelihood-of-next-mention study reveals a pattern in the predicted direction: when Goal referents were also Subjects they were rated as more likely to be referred to than when they were non-Subjects (172 versus 154) and similarly, Sources were rated as being more likely to be referred to when they were Subjects as opposed to non-Subjects (86 versus 68). There was no significant effect of grammatical Subject-hood, however.

One explanation for this discrepancy between the likelihood of next mention findings in this study and the general pattern for Subject re-mentions (Brennan, 1995; Arnold, 1998 dissertation) is that the predictability for Subject continuations is likely less strong in single sentence contexts as compared to longer discourses. For example, the predictability that *Lady Mannerly* will be referred to again immediately given the sentence “*Lady Mannerly handed a painting of the two of them to Sir Barnes*” is likely lower than after the sentences “*Lady Mannerly was such a thoughtful gift giver. She was known to be very kind and generous with gifts for special occasions. She got such delight from giving people things she knew they'd love. Lady Mannerly handed a painting of the two of them to Sir Barnes*”. In the second set of sentences additional factors are contributing to *Lady Mannerly*'s topicality. She has been repeatedly mentioned and referred to with a pronoun, both of which contribute to her topicality.

The choice to introduce *Lady Mannerly* as the Subject in the first sentence isn't a terribly strong cue that she will continue to be talked about. Even if the speaker actually wanted to continue talking about the painting, or *Sir Barnes*, introducing *Lady Mannerly* as the Subject would be perfectly acceptable, as speakers prefer active constructions, and for animate referents to take the role of Subject. In the final sentence of the second example, *Lady Mannerly* is clearly

the topic of the local discourse, was introduced as the Subject, and was repeatedly referred to in that grammatical position. The continued reference, parallelism between sentences, and topicality are all very strong contributors to the predictability that she will be referred to again. The implication of this comparison is that Subject referents in longer discourses tend to also have other features that contribute to their topicality, in addition to the grammatical role information

Under Centering Theory, in the single-sentence example: “*Lady Mannerly handed a painting of the two of them to Sir Barnes*”, *Lady Mannerly* is the Cp, or preferred center, as she was introduced as the grammatical Subject. She is not, however, the established Cb, or backward-looking center, as she has not been mentioned before. In the second example *Lady Mannerly* is the established Cb, as she was referred to in the Subject role in the preceding sentence. This is a stronger cue that she is the focus of attention and should be pronominalized upon repeated mention.

Returning to the previous example, the semantic bias to refer to *Sir Barnes* following this set of sentences: “*Lady Mannerly was such a thoughtful gift giver. She was known to be very kind and generous with gifts for special occasions. She got such delight from giving people things she knew they'd love. Lady Mannerly handed a painting of the two of them to Sir Barnes*” may compete with the topicality bias for *Lady Mannerly*. However, the semantic bias effect for *Sir Barnes* is likely not as strong as in the single sentence example “*Lady Mannerly handed a painting of the two of them to Sir Barnes*”, due to the discourse context. This comparison underscores the conclusion that discourse coherence is crucial for semantic bias effects. Given the discourse context, it isn't surprising that participants in this task did not have strong preferences to rate Subject continuations as being more predictable, although in general data about actual continuations may support the likelihood of Subject continuations.

## Evaluation of Mechanisms

Two mechanisms were proposed to account for the role of predictability in choice of referring expression. Under the referent accessibility mechanism, predictability could play a direct effect on referential accessibility by directing working memory resources to the representations of predictable referents. Topicality would also directly influence the allocation of working memory resources.

Under the facilitation mechanism, predictability could play a less direct role in influencing accessibility. Under this mechanism topicality alone would direct the allocation of working memory resources to certain referents. However, predictability would facilitate other aspects of the production process. If more predictable referents or events sped up production in general, this faster planning and production might allow referents to be retrieved more quickly. Thus, more predictable referents would be retrieved with less time for activation to decay, allowing for selection of a reduced form.

However, the factors that contribute to ease of planning, or signal planning difficulty, did not influence pronoun/zero selection. Disfluency, latency to begin speaking, and ease of verb retrieval were not predictors of referential form, but they were related to each other. Participants began speaking later if they were disfluent, and they began speaking more quickly if the verb was easier to retrieve. These findings would suggest that predictability does not influence reference production via the facilitation mechanism: by generally speeding the production process. However, these results should be interpreted with the caveat that the latency measure used in this experiment was not tightly locked to stimulus onset. Participants were presented with both pictures in a pair at once, and latency was measured from the offset of the description of the first picture. Participants could have begun planning their continuations at different points after

picture presentation. Therefore, this measure of planning time was not tightly linked to when participants were made aware of the coherence relations between the events. A more tightly-locked measure of latency could use eye-tracking to measure latency to begin speaking from participants' first looks to the picture they would be talking about.

### Conclusions and Future Directions

The manipulation of predictability in this set of studies was narrowly focused on one kind of predictability, derived from a particular class of verbs and rooted in the discourse context. The set of studies was focused on the implications of thematic role predictability information for referring expressions. However, there are many other forms of predictability and many other levels of language production where predictability information could play a role. Furthermore, it is important to make the distinction between the role that predictability played in the current experiment and its likely role during normal conversation. In the current experiments predictability was manipulated to encourage continuations about a particular character, or to evaluate how the pre-determined character was referred to. The event-retelling paradigm in particular is most analogous to describing an on-going event, where what you say next is determined by upcoming actions.

What role could predictability play during normal conversation? Imagine that you call your friend to tell her about a job offer you got. You begin the conversation wishing to convey that piece of information, but after you have discussed it you move on to other topics. Wishing to be engaging and trade relevant information, while listening to your friend you may anticipate what she will say next, and think of related information you could add. If she were talking about her engagement you might anticipate that she would want to discuss the bridesmaid dresses, and

your prediction about these related, upcoming topics might lead you to bring up that issue yourself.

Therefore, conversational partners are likely engaged in making predictions constantly during conversation, on many different linguistic levels. You may be anticipating the particular words your friend will use to describe the bridesmaid dresses, and you may also be anticipating more general concepts such as the entertainment at the wedding. Crucially, this predictability likely also influences your own production. Wishing to be a cooperative and relevant conversational partner, you select concepts to discuss that are related to the topics you predict your friend will be introducing.

In conclusion, the current study has found that predictability does play a role in determining referential form. This is an important finding because the role of predictability, while previously reported during comprehension, had not been found to influence production processes. Further work can focus on determining to what extent participants during normal conversation are engaged in predictions, what kinds of predictability participants are sensitive to, and how this predictability information manifests itself in speech, in terms of prosodic choices and form selection.

## APPENDIX A: STIMULI FROM EXPERIMENT 1

### Written Stimuli, List A:

The maid and the butler made their bed  
Sir Barnes received a painting of the two of them from Lady Mannerly  
The chef and the chauffeur brought in some groceries  
The maid accepted some coffee from the butler  
The chef and the chauffeur sharpened some knives  
The butler handed an apron to the maid  
Lady Mannerly saw the maid shine Sir Barnes's shoes  
The butler polished some silver  
Lady Mannerly gave a backrub to Sir Barnes  
Sir Barnes wrote a love letter to the maid  
Lady Mannerly borrowed a picnic basket from the chef  
The chauffeur drove Sir Barnes to town  
Lady Mannerly gave a basket of laundry to the maid  
The chef and Lady Mannerly watered the plants in the kitchen  
Sir Barnes bought earrings from the sales clerk  
The maid held open the house door for Sir Barnes  
Lady Mannerly took some toast from the chef  
The maid dusted pictures  
Lady Mannerly made some sandwiches  
The maid accepted the earrings from Sir Barnes  
Sir Barnes left the love letter on the table  
The maid read her love letter  
Lady Mannerly handed the picnic basket to Sir Barnes  
The butler found the letter in the apron  
The chef handed a cookbook to the maid.  
Lady Mannerly handed the maid a duster and a broom  
The chauffeur took the letter from the butler  
The maid gave Lady Mannerly a glass of champagne  
The rifle salesman sold a gun to the butler  
Sir Barnes and Lady Mannerly read the newspaper  
The butler purchased the bullets from the chauffeur  
Sir Barnes and Lady Mannerly napped in their chairs  
The chef taught the maid how to frost a cake  
The butler and the chauffeur smoked cigars outside  
The maid received a fur coat from Sir Barnes  
The chauffeur gave a silencer to the butler  
The maid got a key from Sir Barnes  
The butler took a flask out of his pocket

The maid handed a piece of cake to Sir Barnes  
The chef and Lady Mannerly planned the week's menu  
The chauffeur taught the butler how to shoot  
Sir Barnes and the butler played some pool  
Sir Barnes threw his pool cue at the butler  
Lady Mannerly took the ruined laundry from the maid  
Lady Mannerly dumped the laundry out onto the floor  
The maid handed the fur coat to the butler  
Sir Barnes kicked the pool table  
The butler got some ice from the chef  
The chef and the chauffeur went to the orchard  
Sir Barnes gave the check to the butler  
The butler ripped up the check  
The chauffeur handed the baskets to the chef  
The chef and the chauffeur saw Sir Barnes on the kitchen floor

#### **Written Stimuli, List B:**

The maid and the butler made their bed  
Lady Mannerly gave a painting of the two of them to Sir Barnes  
The chef and the chauffeur brought in some groceries  
The butler brought some coffee to the maid  
The chef and the chauffeur sharpened some knives  
The maid took an apron from the butler  
Lady Mannerly saw the maid shine Sir Barnes's shoes  
The butler polished some silver  
Sir Barnes got a backrub from Lady Mannerly  
Sir Barnes wrote a love letter to the maid  
The chef loaned a picnic basket to Lady Mannerly  
The chauffeur drove Sir Barnes to town  
The maid took a basket of laundry from Lady Mannerly  
The chef and Lady Mannerly watered the plants in the kitchen  
The sales clerk sold earrings to Sir Barnes  
The maid held open the house door for Sir Barnes  
The chef gave some toast to Lady Mannerly  
The maid dusted pictures  
Lady Mannerly made some sandwiches  
Sir Barnes gave the earrings to the maid  
Sir Barnes left the love letter on the table  
The maid read her love letter  
Sir Barnes took the basket from Lady Mannerly

The butler found the letter in the apron  
The maid took a cookbook from the chef  
Lady Mannerly handed the maid a duster and a broom  
The butler handed the letter to the chauffeur  
The maid gave Lady Mannerly a glass of champagne  
The butler bought a gun from the rifle salesman  
Sir Barnes and Lady Mannerly read the newspaper  
The chauffeur sold bullets to the butler  
Sir Barnes and Lady Mannerly napped in their chairs  
The maid learned from the chef how to frost a cake  
The butler and the chauffeur smoked cigars outside  
Sir Barnes gave a fur coat to the maid  
The chauffeur gave a silencer to the butler  
Sir Barnes gave the key to the maid  
The butler took a flask out of his pocket  
Sir Barnes took a piece of cake from the maid  
The chef and Lady Mannerly planned the week's menu  
The butler learned from the chauffeur how to shoot  
Sir Barnes and the butler played some pool  
Sir Barnes threw his pool cue at the butler  
The maid gave the ruined laundry back to Lady Mannerly  
Lady Mannerly dumped the laundry out onto the floor  
The butler grabbed the fur coat from the maid  
Sir Barnes kicked the pool table  
The chef gave some ice to the butler  
The chef and the chauffeur went to the orchard  
The butler grabbed his check from Sir Barnes  
The butler ripped up the check  
The chef got the baskets from the chauffeur  
The chef and the chauffeur saw Sir Barnes on the kitchen floor

## **APPENDIX B: CLAUSE CODING SCHEMA**

A clause was defined as a main or subordinate independent clause, including all arguments and adjuncts. Subordinate clauses begin with adverbs like because, when, since, while, so. Clauses could be finite or infinite. Clauses with zero subjects still count as separate clauses.

The butler was outside plotting	=main clause
until it was midnight	=subordinate clause
The butler plotted until midnight	= main clause (“until midnight” isn’t a clause – no verb)
The butler plotted outside	=main clause
smoking all day long	=nonfinite subordinate clause
...and smoked a lot	=finite coordinate clause
...instead of doing his work	=not a clause
The butler went to town	=main clause
Purchased some bullets	=main clause with elliptical subject
The butler went to town	=main clause
Then purchased some bullets	=main clause with elliptical subject

Coordinate verbs should be included in the same clause if they are part of the same verb phrase:

The family ate and drank well all evening.

The butler went and got the gun.

But not when the whole sentence is coordinated:

The family ate a big meal

Relative clauses should not be separated-they are included in the same clause as the NP that contains them:

The gun he purchased was shiny.

Clauses that modify the entire previous clause should be counted as a separate clause:

Sir Barnes had purchased the coat at a fur store

Which made Lady Mannerly upset =separate clause

Nonfinite or adjectival clauses that have as their Subject the Object of the higher clause should be included in the higher clause:

The maid is waiting for Sir Barnes to come home

But clauses like “to pick a fight” should be separate:

The maid is waiting for Sir Barnes

To pick a fight =separate clause

Repeated words should be included in a clause if they are replaced by words that don’t start a new clause, e.g.:

<And then um, she, um,> she grabbed the basket

Except if the speaker really seems to break off the clause and start a new one:

And then she grab-

Then she took the basket from him. =separate clause

Other considerations:

In the event of disfluencies or other words that don’t fit neatly within a clause, or seem like they could belong to one or another clause, use prosody to determine which seems correct.

## APPENDIX C: STIMULI FROM EXPERIMENT 2

### Written Stimuli, List A:

The teacher and the secretary made their bed  
Michael received a painting of the two of them from Mary.  
The bookkeeper and the accountant brought in some groceries  
The policeman accepted some coffee from the policewoman  
The chef and the waitress sharpened some knives  
The lawyer handed an apron to the maid  
Susan saw the salesclerk shine David's shoes  
The butler polished some silver  
Linda gave a backrub to James  
Robert wrote a love letter to the model  
Karen borrowed a picnic basket from the caterer  
The chauffeur drove John to town  
Patricia gave a basket of laundry to the house cleaner  
The gardener and Debra watered the plants in the kitchen  
William bought earrings from the jeweler  
The janitor held open the door for Mark  
Cynthia took some toast from the nurse  
The bartender dusted pictures  
Barbara made some sandwiches  
The designer accepted the scarf from Richard  
Thomas left the notebook on the table  
The pharmacist read her mail  
Donna handed the towel to Steven  
The dentist found the book in the drawer  
The coach handed a playbook to the wrestler  
Pamela handed the plumber a wrench and a plunger  
The librarian took the cup from the babysitter  
The server gave Nancy a glass of champagne  
The rifle salesman sold a gun to the firefighter  
Sharon and Joseph read the newspaper  
The midwife purchased the house from the realtor  
Cheryl and Timothy napped in their chairs  
The ballerina taught the cheerleader how to park the car  
The doctor and the veterinarian smoked cigars outside  
The actress received a fur coat from Kenneth  
The electrician gave a bill to the doorman  
The stewardess got a key from Paul

The judge took a flask out of his pocket.  
The professor handed a piece of cake to Kate  
The assistant and Kathy planned the week's menu  
The gambler taught the bouncer how to bet  
Gary and the counselor played some pool  
Donald threw his shoe at the mechanic  
Sandra took the ruined dress from the seamstress  
Brenda dumped the clothes out onto the floor  
The hunter handed the meat to the saleswoman  
Kevin kicked the arcade game  
The hostess got some ice from the grocer  
The farmer and the rancher went to the orchard  
Sally gave the check to the nanny  
The landscaper ripped up the sign  
The housewife handed the bottles to the milkman  
The mayor and the commissioner saw Larry on the dining room floor

#### Written Stimuli, List B:

The teacher and the secretary made their bed.  
Mary gave a painting of the two of them to Michael.  
The bookkeeper and the accountant brought in some groceries.  
The policewoman brought some coffee to the policeman  
The chef and the waitress sharpened some knives  
The maid took an apron from the lawyer  
Susan saw the salesclerk shine David's shoes  
The butler polished some silver  
James got a backrub from Linda  
Robert wrote a love letter to the model  
The caterer loaned a picnic basket to Karen  
The chauffeur drove John to town  
The house cleaner took a basket of laundry from Patricia  
The gardener and Debra watered the plants in the kitchen  
The jeweler sold earrings to William  
The janitor held open the door for Mark  
The nurse gave some toast to Cynthia  
The bartender dusted pictures  
Barbara made some sandwiches  
Richard gave the scarf to the designer  
Thomas left the notebook on the table

The pharmacist read her mail  
Steven took the towel from Donna  
The dentist found the book in the drawer  
The wrestler took a playbook from the coach  
Pamela handed the plumber a wrench and a plunger  
The babysitter handed the cup to the librarian  
The server gave Nancy a glass of champagne  
The firefighter bought a gun from the rifle salesman  
Sharon and Joseph read the newspaper  
The realtor sold the house to the midwife  
Cheryl and Timothy napped in their chairs  
The cheerleader learned from the ballerina how to park the car  
The doctor and the veterinarian smoked cigars outside  
Kenneth gave a fur coat to the actress  
The electrician gave a bill to the doorman  
Paul gave the key to the stewardess  
The judge took a flask out of his pocket.  
Kate took a piece of cake from the professor  
The assistant and Kathy planned the week's menu  
The bouncer learned from the gambler how to bet  
Gary and the counselor played some pool  
Donald threw his shoe at the mechanic  
The seamstress gave the ruined dress back to Sandra  
Brenda dumped the clothes out onto the floor  
The saleswoman grabbed the meat from the hunter  
Kevin kicked the arcade game  
The grocer gave some ice to the hostess  
The farmer and the rancher went to the orchard  
The nanny grabbed her check from Sally  
The landscaper ripped up the sign  
The milkman got the bottles from the housewife  
The mayor and the commissioner saw Larry on the dining room floor

## APPENDIX D: STIMULI FROM EXPERIMENT 3

### Written stimuli, List A:

The maid and the butler made their bed.  
Sir Barnes received a painting of the two of them from Lady Mannerly.  
The chef and the chauffeur brought in some groceries.  
The maid accepted some coffee from the butler.  
The chef and the chauffeur sharpened some knives.  
The butler handed an apron to the maid.  
Lady Mannerly saw the maid shine Sir Barnes's shoes.  
The butler polished some silver.  
Lady Mannerly gave a backrub to Sir Barnes.  
Sir Barnes wrote a love letter to the maid.  
Lady Mannerly borrowed a picnic basket from the chef.  
The chauffeur drove Sir Barnes to town.  
Lady Mannerly gave a basket of laundry to the maid.  
The chef and Lady Mannerly watered the plants in the kitchen.  
Sir Barnes bought earrings from a sales clerk.  
The maid held open the house door for Sir Barnes.  
Lady Mannerly took some toast from the chef.  
The maid dusted pictures.  
Lady Mannerly made some sandwiches.  
The maid accepted the earrings from Sir Barnes.  
Sir Barnes left the love letter on the table.  
The maid read her love letter.  
Lady Mannerly handed the picnic basket to Sir Barnes.  
The butler found the letter in the apron.  
The chef handed a cookbook to the maid.  
Lady Mannerly handed the maid a duster and a broom.  
The chauffeur took the letter from the butler.  
The maid gave Lady Mannerly a glass of champagne.  
The rifle salesman sold a gun to the butler.  
Sir Barnes and Lady Mannerly read the newspaper.  
The butler purchased the bullets from the chauffeur.  
Sir Barnes and Lady Mannerly napped in their chairs.  
The chef taught the maid how to frost a cake.  
The butler and the chauffeur smoked cigars outside.  
The maid received a fur coat from Sir Barnes.  
The chauffeur gave a silencer to the butler.  
The maid got a key from Sir Barnes.  
The butler took a flask out of his pocket.

The maid handed a piece of cake to Sir Barnes.  
The chef and Lady Mannerly planned the week's menu.  
The chauffeur taught the butler how to shoot.  
Sir Barnes and the butler played some pool.  
Sir Barnes threw his pool cue at the butler.  
Lady Mannerly took the ruined laundry from the maid.  
Lady Mannerly dumped the laundry out onto the floor.  
The maid handed the fur coat to the butler.  
Sir Barnes kicked the pool table.  
The butler got some ice from the chef.  
The chef and the chauffeur went to the orchard.  
Sir Barnes gave the check to the butler.  
The butler ripped up the check.  
The chauffeur handed the baskets to the chef.  
The chef and chauffeur saw Sir Barnes on the kitchen floor.

**Written Stimuli, List B:**

The maid and the butler made their bed.  
Lady Mannerly gave a painting of the two of them to Sir Barnes.  
The chef and the chauffeur brought in some groceries.  
The butler brought some coffee to the maid.  
The chef and chauffeur sharpened some knives.  
The maid took an apron from the butler.  
Lady Mannerly saw the maid shine Sir Barnes's shoes.  
The butler polished some silver.  
Sir Barnes got a backrub from Lady Mannerly.  
Sir Barnes wrote a love letter to the maid.  
The chef loaned a picnic basket to Lady Mannerly.  
The chauffeur drove Sir Barnes to town.  
The maid took a basket of laundry from Lady Mannerly.  
The chef and Lady Mannerly watered the plants in the kitchen.  
The sales clerk sold earrings to Sir Barnes.  
The maid held open the house door for Sir Barnes.  
The chef gave some toast to Lady Mannerly.  
The maid dusted pictures.  
Lady Mannerly made some sandwiches.  
Sir Barnes gave the earrings to the maid.  
Sir Barnes left the love letter on the table.  
The maid read her love letter.  
Sir Barnes took the picnic basket from Lady Mannerly.  
The butler found the letter in the apron.  
The maid took a cookbook from the chef.

Lady Mannerly handed the maid a duster and a broom.  
The butler handed the letter to the chauffeur.  
The maid gave Lady Mannerly a glass of champagne.  
The butler bought a gun from the rifle salesman.  
Sir Barnes and Lady Mannerly read the newspaper.  
The chauffeur sold bullets to the butler.  
Sir Barnes and Lady Mannerly napped in their chairs.  
The maid learned from the chef how to frost a cake.  
The butler and the chauffeur smoked cigars outside.  
Sir Barnes gave a fur coat to the maid.  
The chauffeur gave a silencer to the butler.  
Sir Barnes gave the key to the maid.  
The butler took a flask out of his pocket.  
Sir Barnes took a piece of cake from the maid.  
The chef and Lady Mannerly planned the week's menu.  
The butler learned from the chauffeur how to shoot.  
Sir Barnes and the butler played some pool.  
Sir Barnes threw his pool cue at the butler.  
The maid gave the ruined laundry back to Lady Mannerly.  
Lady Mannerly dumped the laundry out onto the floor.  
The butler grabbed the fur coat from the maid.  
Sir Barnes kicked the pool table.  
The chef gave some ice to the butler.  
The chef and the chauffeur went to the orchard.  
The butler grabbed the check from Sir Barnes.  
The butler ripped up the check.  
The chef got the baskets from the chauffeur.  
The chef and chauffeur saw Sir Barnes on the kitchen floor.

## APPENDIX E: CONTINUATION CODING SCHEMA FROM EXPERIMENT 1(ROHDE)

### INTRO:

There are 7 categories for annotation

- Elaboration (elab)
- Explanation (exp)
- Occasion (occ)
- Parallel (par)
- Result (res)
- Violated Expectation (v)
- Background (back)

Any pair of sentences can be related by one or more of the 7 categories. Your Goal is to identify the intent of the author (or your best guess at it) and to pick the most likely coherence relation(s). If there is more than one verb or more than one event in the second sentence, just use the matrix verb/first event. Below are some general guidelines and examples.

### ELABORATION:

- elaborate on the same event -- how it is carried out or where/when
  - ex. John hit Bob. John hit him hard in the nose
  - ex. John played piano... He played Mozart
  - ex. John made a sandwich.. John made it with a knife

### EXPLANATION:

- explanation about the previous event or general information about the cause of an event
  - ex. John surprised Bill...He bought him flowers.
  - ex:John aggravated Susan...He pulled her hair.
  - ex:John bored Bill...He told endless stories.

### OCCASION:

- temporal relation between two sentences where second sentence describes an event that follows the first sentence. **If the two events are causally related, use RESULT instead.**
  - ex. john hit bob. next john went to the store
  - ex. john threw a ball to bob. he caught it.
  - ex. john cooked dinner. afterwards he did the dishes.
- could be marked by then/later/afterwards/next

### RESULT:

- causal result of previous event
- information about someone's reaction or resulting emotion
  - ex. john hit bob. bob got really mad.
  - ex. john gave a book to mary. she said thanks.
- could be marked by "as a result"

## PARALLEL:

- things to look for...
- same verb/different referents
- same referents, different verb
  - ex. john hates mary. mary hates john.
- could be marked by "similarly"

## VIOLATED-EXPECTATION:

- an unexpected outcome given general real-world knowledge about likely events and their typical consequences/reactions
  - ex. john threw a ball to mary. she dropped it.
  - ex. john insulted mary. she was not offended.
- often good with 'but'/'however'

BACKGROUND: a more general category for continuations that contain background info that elaborates on some aspect of the event (more canonical 'Elaborations' retell the same event with additional information, but we often see cases of additional information that picks up on only a subpart of the event and gives more details about that particular part or particular referent) --> these are often collapsed with 'Elaboration' for analysis

- ex. played the piano... been playing for 5 years
- ex. went to the library.. they often go to the library
- simultaneous
  - ex. John waited to see Mary. He didn't want her to think he was eager.
  - ex. John was gardening. He was thinking about what to plant next year.
- extraneous details about one of the referents
  - ex. Charlotte stood next to Parker. She has seven sisters.
- prior state of one of the referents
  - ex. John frightened Mary. He had been planning to do so all day.

---

## Notes on deciding between categories:

In general, assign a causal reading if it's available (even if a non-causal relation also holds)

### Occasion vs. Result

- ex. John gave a book to Bob. Bob said thanks.
  - ... thanks as what happened next or as a result

-> Assign causal Result if available

### Elaboration vs. Result

- ex. David was amused by Josh. He laughed at Josh's many jokes.
  - ... laughter as result of amusement or elaboration?
- ex. Holly was astonished by Caitlin. She had to practically pick her jaw up off of the floor.
  - ... jaw drop as result of astonishment or elaboration?

-> Assign causal reading (Result) if available

### Explanation vs. Elaboration

ex. Tony disappointed Brandon. He cheated on his test, and lied about it.

...disappointed by cheating or because of cheating?

... this would be classed as explanation of the disappointment

ex. James charmed Bob. He really wanted the job and tried his best to impress Bob.

... this would be classed as an explanation of why James turned on the charm

ex. Chad humiliated Kyle. He did this daily,

... this would be classed as an elaboration of when/where

### **In general, assign causal reading (Explanation) if available given the following:**

(a) in previous work we have seen continuations like this when we prompt

participants with 'because' (Heather offended Bob because ... she called him fat)

(b) participants in this experiment produce the connective 'because'

which signals that they (or some of them at least) infer an Explanation reading (Brandon was disappointed by Tony. ... because tony cheated on a test)

(c) similar continuations act like Explanations when the context

sentence is passive:

(i) Mike aggravated Christopher. Mike was always making stupid jokes

(ii) Christopher was aggravated by Mike. Mike was being a jerk

## **APPENDIX F: TRANSCRIPT OF BACKGROUND VIDEO FOR EXPERIMENT 4, 5, &6**

You are a tabloid photographer, and you were assigned to cover the manor of Sir Barnes and Lady Mannerly, who are very wealthy British people who are notoriously secretive. Your assignment was to get photographs of them and find out as much as possible about them to write an article. They do not have children, but they do have help who live in the manor as well, a butler and a maid, who are married, as well as a chef and a chauffeur. The butler is responsible for the day-to-day operations of the house, and is friends with the chauffeur. The chauffeur does the driving and keeps the automobiles maintained. The maid does the cleaning and helps serve the food. The chef is a motherly figure. Here is a diagram of the house, which you may find helpful. Under the guise of being a long-lost cousin, you were a guest at their manor over the weekend. Unbeknownst to them, you installed cameras all over the house and on their clothing, so you captured their every move. Shockingly, a murder occurred while you were there. The pictures from your cameras, which you haven't even seen yet, have been taken as evidence to help solve the crime. You are going to review the photographs with a detective to help us piece together the crime. The photographs were time stamped originally, so you will view them in the order in which they were taken. (End of background video for Experiments 5 & 6)

(Continuation of background video for Experiment 4)

The detective knows some of the events from that weekend, but not everything. So the detective will describe the first event in each scenario and you will tell us, from the photo, what happened next. Now you will have the opportunity to see your photos, in order, before talking about them with the detective.

## **APPENDIX G: POST-EXPERIMENT QUESTIONNAIRE FROM EXPERIMENT 4**

Who was murdered?

Why was he murdered?

Who do you think committed the crime?

With what weapon was the crime committed?

Are you familiar with the game Clue?

Yes   No

Have you played the Clue game?

Yes   No

Have you watched the Clue movie?

Yes   No

What do you think this experiment was about?

Do you have any other comments about it?

## APPENDIX H: STIMULI FROM EXPERIMENT 5

### Written stimuli, List A:

The maid and the butler made their bed  
Sir Barnes received a painting of the two of them from Lady Mannerly  
The chef and the chauffeur brought in some groceries.  
The maid accepted some coffee from the butler.  
The chef and the chauffeur sharpened some knives.  
The butler handed an apron to the maid.

Lady Mannerly saw the maid shine Sir Barnes's shoes.  
The butler polished some silver.  
Lady Mannerly gave a backrub to Sir Barnes.  
Sir Barnes wrote a love letter to the maid.  
Lady Mannerly borrowed a picnic basket from the chef.  
The chauffeur drove Sir Barnes to town.  
Lady Mannerly gave a basket of laundry to the maid.  
The chef and Lady Mannerly watered the plants in the kitchen.  
Sir Barnes bought earrings from a sales clerk.  
The maid held open the house door for Sir Barnes.  
Lady Mannerly took some toast from the chef.  
The maid dusted pictures.

Lady Mannerly made some sandwiches.  
The maid accepted the earrings from Sir Barnes.  
Sir Barnes left the love letter on the table.  
The maid read her love letter.  
Lady Mannerly handed the picnic basket to Sir Barnes.  
The butler found the letter in the apron.  
The chef handed a cookbook to the maid.  
Lady Mannerly handed the maid a duster and a broom.  
The chauffeur took the letter from the butler.  
The maid gave Lady Mannerly a glass of champagne.  
The rifle salesman sold a gun to the butler.  
Sir Barnes and Lady Mannerly read the newspaper.  
The butler purchased the bullets from the chauffeur.  
Sir Barnes and Lady Mannerly napped in their chairs.  
The chef taught cake decorating to the maid.  
The butler and the chauffeur smoked cigars outside.  
The maid received a fur coat from Sir Barnes.  
The chauffeur gave a silencer to the butler.  
The maid got a key from Sir Barnes.  
The butler took a flask out of his pocket.  
The maid handed a piece of cake to Sir Barnes.  
The chef and Lady Mannerly planned the week's menu.  
The chauffeur taught shooting techniques to the butler.  
Sir Barnes and the butler played some pool.  
Sir Barnes threw his pool cue at the butler.  
Lady Mannerly took the ruined laundry from the maid.

The maid opened the curtains  
Sir Barnes threw the painting in a closet  
The chef unpacked the groceries  
The butler lit a fire  
The chauffeur hid one in a drawer  
The butler tied the ribbon on it  
Lady Mannerly saw Sir Barnes kiss the maid on the hand  
The butler put a candlestick in his pocket  
Sir Barnes stormed out.  
Sir Barnes put the letter in an envelope  
Lady Mannerly folded a picnic blanket.  
Sir Barnes went into a jewelry store.  
The maid poured bleach on the laundry

The chef patted Lady Mannerly on the back  
The sales clerk wrapped the gift.  
Sir Barnes whispered in her ear  
Lady Mannerly pushed her plate away  
The butler saw the maid kiss a picture of Sir Barnes  
Lady Mannerly packed the sandwiches in the picnic basket  
The maid exclaimed in happiness.  
The maid found the love letter on the table  
The maid put it in her apron pocket  
Sir Barnes threw the picnic basket in the corner.  
The butler opened the letter  
The chef pointed to a cake recipe in the book.  
The maid threw them on the ground  
The butler pounded his hand into his fist  
Lady Mannerly threw it in her face  
The butler polished the gun  
Sir Barnes checked his watch  
The butler loaded the bullets in the gun.  
Sir Barnes snuck out of the library  
The maid drew a heart on the top of the cake  
They shook hands  
Sir Barnes closed the curtains.  
The butler screwed it onto the end of the gun  
The maid locked the door  
The butler took a drink  
The maid pointed at the heart on top of the cake  
The chef turned the radio on  
The chauffeur wiped off the gun  
The butler won the game  
Sir Barnes and the butler fought  
The maid laughed

Lady Mannerly dumped the laundry out onto the floor.  
The maid handed the fur coat to the butler.  
Sir Barnes kicked the pool table.  
The butler got some ice from the chef.  
The chef and the chauffeur went to the orchard.  
Sir Barnes gave the check to the butler.  
The butler ripped up the check.  
The chauffeur handed the baskets to the chef.  
The chef and chauffeur saw Sir Barnes on the kitchen floor.

Lady Mannerly sat down and cried  
The butler looked at the label  
Sir Barnes broke a pool cue  
The chef wrung her hands  
The chauffeur checked his watch  
Sir Barnes laughed at him  
The butler reached for the gun  
The chauffeur opened the door to the house  
  
The chef picked up the gun from the floor

### **Written stimuli, List B:**

The maid and the butler made their bed.  
Lady Mannerly gave a painting of the two of them to Sir Barnes.  
  
The chef and the chauffeur brought in some groceries.  
The butler brought some coffee to the maid.  
The chef and chauffeur sharpened some knives.  
The maid took an apron from the butler.  
  
Lady Mannerly saw the maid shine Sir Barnes's shoes.  
The butler polished some silver.  
Sir Barnes got a backrub from Lady Mannerly.  
Sir Barnes wrote a love letter to the maid.  
The chef loaned a picnic basket to Lady Mannerly.  
The chauffeur drove Sir Barnes to town.  
The maid took a basket of laundry from Lady Mannerly.  
The chef and Lady Mannerly watered the plants in the kitchen.  
  
The sales clerk sold earrings to Sir Barnes.  
The maid held open the house door for Sir Barnes.  
The chef gave some toast to Lady Mannerly.  
The maid dusted pictures.  
  
Lady Mannerly made some sandwiches.  
Sir Barnes gave the earrings to the maid.  
Sir Barnes left the love letter on the table.  
The maid read her love letter.  
Sir Barnes took the picnic basket from Lady Mannerly.  
The butler found the letter in the apron.  
The maid took a cookbook from the chef.  
Lady Mannerly handed the maid a duster and a broom.

The maid opened the curtains  
  
Sir Barnes put the painting in a closet  
The chef unpacked the groceries  
He lit a fire  
The chauffeur hid one in a drawer  
The butler tied the ribbon on it  
Lady Mannerly saw Sir Barnes kiss the maid on the hand  
  
The butler put a candlestick in his pocket  
Sir Barnes stormed out.  
Sir Barnes put the letter in an envelope  
Lady Mannerly folded a picnic blanket.  
Sir Barnes went into a jewelry store.  
  
The maid poured bleach on the laundry  
  
The chef patted Lady Mannerly on the back  
The sales clerk wrapped the gift.  
Sir Barnes whispered in her ear  
Lady Mannerly pushed her plate away  
The butler saw the maid kiss a picture of Sir Barnes  
Lady Mannerly packed the sandwiches in the picnic basket  
  
The maid exclaimed in happiness.  
The maid found the love letter on the table  
The maid put it in her apron pocket  
Sir Barnes threw the picnic basket in the corner.  
The butler opened the letter  
The chef pointed to a cake recipe in the book.  
The maid threw them on the ground

The butler handed the letter to the chauffeur.  
The maid gave Lady Mannerly a glass of champagne.  
The butler bought a gun from the rifle salesman.  
Sir Barnes and Lady Mannerly read the newspaper.  
The chauffeur sold bullets to the butler.  
Sir Barnes and Lady Mannerly napped in their chairs.  
The maid learned cake decorating from the chef  
The butler and the chauffeur smoked cigars outside.  
Sir Barnes gave a fur coat to the maid.  
The chauffeur gave a silencer to the butler.  
Sir Barnes gave the key to the maid.  
The butler took a flask out of his pocket.  
Sir Barnes took a piece of cake from the maid.  
The chef and Lady Mannerly planned the week's menu.  
The butler learned shooting techniques from the chauffeur  
Sir Barnes and the butler played some pool.  
Sir Barnes threw his pool cue at the butler.  
The maid gave the ruined laundry back to Lady Mannerly.  
Lady Mannerly dumped the laundry out onto the floor.  
The butler grabbed the fur coat from the maid.  
Sir Barnes kicked the pool table.  
The chef gave some ice to the butler.  
The chef and the chauffeur went to the orchard.  
The butler grabbed the check from Sir Barnes.  
The butler ripped up the check.  
The chef got the baskets from the chauffeur.  
The chef and chauffeur saw Sir Barnes on the kitchen floor.

The butler pounded his hand into his fist  
Lady Mannerly threw it in her face  
The butler polished the gun  
Sir Barnes checked his watch  
The butler loaded the bullets in the gun.  
Sir Barnes snuck out of the library  
The maid drew a heart on the top of the cake  
They shook hands  
Sir Barnes closed the curtains.  
The butler screwed it onto the end of the gun  
The maid locked the door  
The butler took a drink  
The maid pointed at the heart on top of the cake  
The chef turned the radio on  
The chauffeur wiped off the gun  
The butler won the game  
Sir Barnes and the butler fought  
The maid laughed  
Lady Mannerly sat down and cried  
The butler looked at the label  
Sir Barnes broke a pool cue  
The chef wrung her hands  
The chauffeur checked his watch  
Sir Barnes laughed at him  
The butler reached for the gun  
The chauffeur opened the door to the house  
The chef picked up the gun from the floor

## REFERENCES

- Ariel, M. (1990). *Accessing noun-phrase antecedents*. London: Routledge.
- Arnold, J. E. (1998). *Reference Form and Discourse Patterns*. Dissertation, Stanford University.
- Arnold, J. E. (2001). The effect of thematic role on pronoun use and frequency of reference continuation. *Discourse Processes*, 31, 137-162.
- Arnold, J.E. (2008). Reference Production: Production-internal and Addressee-oriented Processes. *Language and Cognitive Processes*. 23(4), 495-527.
- Arnold, J.E. (2010). How speakers refer: the role of accessibility. *Language and Linguistic Compass*, 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., & Griffin, Z. (2007). The Effect of Additional Characters on Choice of Referring Expression: Everyone Competes. *Journal of Memory and Language*. 56(4), 521-536.
- Arnold, J.E., Kahn, J. & Pancani, G. (2012). Audience Design Affects Acoustic Reduction Via Production Facilitation. *Psychological Bulletin and Review*, 19, 505-512
- Arnold, J.E., & Lao, S.C. (in press). Effects of non-shared attention on pronoun comprehension. *Language, Cognition, & Neuroscience*.
- Arnold, J.E., & Nozari, N. (under review). The effects of utterance planning and stimulation of left prefrontal cortex on the production of referential expressions.
- Aylett, M., Turk, A. (2004). The Smooth Signal Redundancy Hypothesis: A Functional Explanation for Relationships between Redundancy, Prosodic Prominence and Duration in Spontaneous Speech. *Language and Speech*. 47(1), 31-56
- Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., & Gildea, D. (1999). Forms of English function words – Effects of disfluencies, turn position, age and sex, and predictability. In Proceedings of ICPHS-99.
- 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.
- Brennan, S.E. (1995). Centering attention in discourse. *Language and Cognitive Processes*, 10, 137-167.

- Brennan, S. E., Friedman, M. W., & Pollard, C. J. (1987). A centering approach to pronouns. *Proceedings, 25th Annual Meeting of the ACL* pp. 155-162. Stanford, CA: Association of Computational Linguistics.
- Buschkuehl, M., Garcia, L. H., Jaeggi, S. M., Bernard, J. A., & Jonides, J. (2014). Neural Effects of Short-Term Training on Working Memory. *Cognitive, Affective & Behavioral Neuroscience*, 14(1), 147–160.
- Chafe, W. (1976). Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Point of View. In Charles N. Li, ed., *Subject and Topic*. New York: Academic Press.
- Chafe, W. (1994). *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: University of Chicago Press.
- Clark, H.H. & Clark, E.V. (1977). *Psychology and Language: An Introduction to Psycholinguistics*. New York: Harcourt, Brace & Jovanovich.
- Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In R. O. Freedle (Ed.), *Discourse production and comprehension* (pp. 1-40). Hillsdale, NJ: Erlbaum.
- Dell, G.S. (1987). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3), 283-321.
- Ferretti, T.R., Kutas, M., McRae, K. (2007). Verb aspect and the activation of event knowledge. *Journal of Experimental Psychology, Learning, Memory and Cognition*, 33(1), 182-196.
- Filmer, H.L., Mattingly, J.B., & Dux, P.E. (2015). Object substitution masking for an attended and foveated target. *Journal of Experimental Psychology: Human Perception and Performance*. 41(1), 6-10.
- 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.
- Frank, A. & Jaeger, T.F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. *Proceedings of the Annual Meeting of the Cognitive Science Society*.
- Fukumura, K., Hyönä, J., & Scholfield, M. (2013). Gender affects semantic competition: the effect of gender in a non-gender marking language. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 39, 1012-1021
- 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.

- Fukumura, K., Van Gompel, R.P.G., & Pickering, M. J. (2010). The use of visual context during the production of referring expressions. *Quarterly Journal of Experimental Psychology*, 63, 1700–1715.
- 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. and Gernsbacher, M.A. (1996). The locus of implicit causality effects in comprehension. *Journal of Memory and Language*, 4, 517-543.
- Gibson, J.W., Gruner, C.R., Kibler, R.J., Kelly, F.J. (1966). A quantitative examination of differences and similarities in written and spoken messages. *Speech Monographs*, 33, 444--451.
- Givon, T. (1983). Topic continuity in discourse: An introduction. In T. Givon (Ed.), *Topic continuity in discourse: A quantitative cross-language study*. (pp. 1-41). Amsterdam: John Benjamins.
- Greene, S.B., & McKoon, G. (1995). Telling something we can't know: Experimental approaches to verbs exhibiting implicit causality. *Psychological Science*, 6, 262-270.
- Gregory, M.L., Raymond, W. D., Bell, A., Fosler-Lussier, E., and Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. In CLS-99, Chicago. University of Chicago.
- Grosz, B.J., Joshi. A.K., Weinstein, S. (1995). Centering: a framework for modeling the local coherence of discourse. *Computational Linguistics*, 21, 203-225.
- Gundel, J. K. (1988). Universals of topic-comment structure In M. Hammond, E. Moravczik and J. Wirth (eds.), *Studies in Syntactic Typology*. (pp. 209-239). Amsterdam:John Benjamins,
- Gundel, J.K., Hedberg, N., Zacharski, R. (1983). Cognitive status and the form of anaphoric expressions in discourse. *Language*, 69, 274-307.
- Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *Proceedings of the 2nd Conference of the North American Chapter of the Association for Computational Linguistics*, Pittsburgh, PA.
- Halliday, M.A.K. (2007 (1979)). *Differences between spoken and written languages: Some implications for literacy teaching*. Language and Education. Volume 9 in the Collected Works of M.A.K. Halliday. Ed J. Webster. London: Continuum.

- Harrell, L. E., Jr. (1957). *A comparison of oral and written language in school age children.* (Monographs of the Society for Research in Child Development 22, Serial number 66.) Lafayette, IN: Child Development Publications.
- Horowitz, M.U., Newman, J.B. (1964). Spoken and written expression: An experimental analysis. *Journal of Abnormal and Social Psychology*, 68, 640--647.
- Jaeger, T. F. (2006). Redundancy and Syntactic Reduction in Spontaneous Speech. PhD thesis, Stanford University, Stanford, CA
- Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, 20, 137–194.
- Jurafsky, D., Bell, A., Fosler-Lussier, E., Girand, C., & Raymond, W. D. (1998). Reduction of English function words in Switchboard. *Proceedings of ICSLP-98*, Sydney.
- 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.
- 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. .
- 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*.
- Lindblom, B. (1990). Explaining phonetic variation: a sketch of the h & h theory. In Hardcastle, W.J. & Marchal, A. (Eds), *Speech Production and Speech Modeling* (pp. 403–439). New York: Springer.

- McDonald, J. L., & MacWhinney, B. J. (1995). The time course of anaphor resolution: Effects of implicit verb causality and gender. *Journal of Memory and Language*, 34, 543-566.
- Nida, E. A. (1967). Linguistic dimensions of literacy and literature. In Shacklock, F. (Ed.). *World literacy manual* (pp. 142-161). New York: Committee on World Literacy and Christian Literature.
- 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., and 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.
- Prince, E.F. (1981). Toward a Taxonomy of Given-New Information. In P. Cole (Ed.), *Radical Pragmatics*, (pp. 233–255). New York: Academic Press.
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42, 107-142.
- Rohde, H., Kehler, A. (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition, and Neuroscience*, 29(8), 1-16.
- Rohde, H. (2008). *Coherence-driven effects in sentence and discourse processing*. Dissertation: UCSD.
- Sartre, J.P. (1964). *The Words*. Gallimard: Paris.
- Seidenberg, M.S. & MacDonald, M.C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, 23, 569-588.
- Son, R.J.J.H, & van Santen, J.P.H. (2005). Duration and spectral balance of intervocalic consonants: A case for efficient communication. *Speech Communication*, 47, 100-123.
- 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.
- Stevenson, R.J., Knott, A., Oberlander, J., & McDonald, S. (2000). Interpreting pronouns and connectives: Interactions amount focusing, thematic roles, and coherence relations. *Language and Cognitive Processes*, 15, 225-262.

- Tatone, D., Hernik, M., Csibra, G. (2015). Giving -- but not taking -- actions afford object feature encoding in absence of outcome selectivity in 12-month-old infants. *Poster presented at the Budapest CEU Conference on Cognitive Development*, Budapest.
- 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 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: Universitätsverlag der TU Berlin.
- Vygotsky, L.S. (1962). *Thought and Language*. Cambridge, MA: MIT Press. (Original work published in 1934).
- Zerkle, S., Rosa, E.C., Arnold, J.E. (2015). *Do addressee gestures influence the effects of predictability on spoken reference form?* Poster to be presented at the CUNY 2015 conference, Los Angeles, CA.
- Zipf, G. (1936). *The Psychobiology of Language*. London: Routledge.