RUNNING HEAD: Thematic roles can affect reference form

Predictability affects production: Thematic roles can affect reference form selection

Elise C. Rosa and Jennifer E. Arnold

University of North Carolina at Chapel Hill

Key words: pronouns, reference form, utterance production, predictability, thematic roles

Address all correspondence to: Jennifer E. Arnold, UNC Chapel Hill, Dept. of Psychology, Davie Hall #337B, CB #3270, Chapel Hill, NC 27599-3270, Email: jarnold@email.unc.edu

ABSTRACT

Speakers use pronouns when referring to information that is topical, recently mentioned, or salient in the discourse. Although such information is often predictable, there is conflicting evidence about whether predictability affects reference form production. This debate centers on the question of whether reference form is influenced by the predictability of certain thematic roles. While some (Arnold, 2001) argue that referents in certain thematic roles are more likely to be pronominalized, others (Fukumura & van Gompel 2010; Rohde & Kehler, 2014) argue that predictability does not play a role in determining referential form. We tested this puzzle in three experiments, using both a richly contextualized production paradigm, and two versions of the standard story-completion paradigm. In all experiments we manipulated the predictability of pairs of characters using transfer verbs. In all three experiments, we found that speakers used more pronouns when talking about the goal referent as opposed to the source. A rating experiment revealed that participants also expect goals to be mentioned more than sources. These results show that thematic role does affect both perceived predictability and the speaker's choice of reference form.

The selection of an appropriate referring expression is an important component of successful communication. For example, in relating a story about a villain, you need to make multiple decisions about how to refer to him. You would likely use a descriptive expression such as *Bob* or *this creepy guy* upon first mention, and when referring to him again, might choose a more reduced expression such as "*he*".

It is well established that speakers use reduced forms (pronouns) under particular discourse conditions, such as when the referent has been recently mentioned, or was in the grammatical subject position of the last sentence (Ariel, 1990, 2001; Arnold, 1998, 2008, 2010; Brennan, 1995; Givon, 1983; Gundel, Hedberg, & Zacharski, 1993). One hypothesis is that recently and prominently mentioned things tend to be topical to the current discourse segment, and that pronouns are selected on the basis of the topicality of the referent (e.g., Givon, 1983; Kehler, Kertz, Rohde, Elman, 2008; Kehler & Rohde, 2013; van Rij, van Rijn, & Hendriks, 2012).

Yet scholars disagree about whether pronoun production is also influenced by semantic considerations. This debate concerns two inter-related questions: 1) Do speakers use pronouns more for entities that are predictable? 2) Do speakers use pronouns more to refer to entities that occurred in certain thematic roles in the previous utterance? These questions are relevant to building a comprehensive model of reference production, and yet the literature includes conflicting claims for both. These questions are linked, because thematic roles are associated with the likelihood that an entity will be re-mentioned in the next sentence.

Given a particular discourse context, comprehenders have consistent expectations that some characters are more likely to be mentioned again, meaning that they are relatively more predictable as referents. For example, the sentences in (1) depict events in which people tend to

assume that one participant is the more likely cause of the event (e.g., Brown & Fish, 1983;

Hartshorne, O'Donnell, & Tenenbaum, 2015). If a causal statement includes a pronoun

("because he…"), participants tend to interpret the pronoun as coreferential with the implicit

cause (Stevenson, Crawley, Kleinman, 1994). Similarly, the sentences in (2) depict transfer-of-

possession events, in which readers tend to expect that a subsequent event will mention the

receiver of the object (Rohde & Kehler, 2014; Stevenson et al., 1994).

1a). The butler blamed the chauffeur because he…. (murdered someone).

1b). The butler impressed the chauffeur because he… (figured out the case).

2a). The butler gave the threatening note to the chauffeur and he… (turned it in to the police).

2b). The butler received a ticking bomb from the chauffeur and he … (chucked it into the river).

In discourses like these, the predictability of a referent being mentioned is identified with its

thematic role in the event. The thematic role is determined by the verb, and represents the

semantic role of the participants in an event. In 1, the stimulus role is the expected continuation

(the chauffeur in 1a, the butler in 1b), while the experiencer is not. In 2, the goal is the expected

continuation (the chauffeur in 1a, the butler in 1b), while the source is not.

  Critically, the effects of thematic roles on referential predictability are closely tied to the

relationship between the two utterances (Ehrlich, 1980; Kehler, 2002; Kehler & Rohde, 2013;

Stevenson et al., 2004). In the implicit causality sentences, people expect the causal character to

be mentioned if they expect the speaker to produce an utterance about the cause of the first event.

This expectation is created by the connector *because* in (1), but if the sentence continues with *so*

he…, expectations can sometimes reverse (Ehrlich, 1980; Stevenson et al., 1994), although it depends on the verb (Hartshorne, et al., 2015). In (2), the expectation of the goal reference is conditioned on the expectation that the speaker will describe the result of the first event (Stevenson et al., 1994).

The question we are concerned with here is what speakers do in production. Following sentences like (1) or (2), does the speaker choose the pronoun *he* more often for predictable thematic roles than others? Critically, this question is debated, and there is conflicting evidence in the literature. In fact, researchers debate both the question of whether thematic roles matter in particular, and whether predictability matters in general.

One view is that thematic roles do influence pronoun production, and that the reason they do is that they modulate referential predictability. Arnold (1998, 2001) proposed that entities become more accessible when they have a high likelihood of being mentioned again in the discourse (see also Givon, 1983; Tily & Piantadosi, 2009), which increases the speaker's likelihood of using pronouns. Arnold's Expectancy Hypothesis suggests that predictability comes from numerous sources, including the fact that grammatical subjects are more likely to be mentioned again than nonsubjects, and that recently mentioned entities are more likely to be mentioned than less recent entities (Arnold, 1998; 2010). In support of this, Arnold (2001) presented results from a story-continuation experiment, in which participants were asked to invent continuations for passages. These passages included a critical transfer-of-possession prompt, e.g. *Lisa gave the leftover pie to Brendan*. Results revealed that when participants referred to the second character, they used pronouns more often for goals than sources. A corpus analysis confirmed that goals are more likely to be mentioned again than sources.

Kaiser, Li, and Holsinger (2011) also report that thematic roles influence pronoun usage, using prompts with agent and patient roles, such as *Mary slapped Lisa…*, and *Lisa was slapped by Mary…*. They found that when speakers chose pronouns, they were more likely to refer to the character in the patient role (Lisa) than the one in the agent role.[1] However, the authors also argue that thematic roles are not linked to predictability, based on the observation that the strong tendency for pronouns to refer to patients in their data was not mirrored by an equally strong preference to continue talking about the patient.

By contrast, several studies have reported the opposite, that thematic roles do not influence the speaker's choice of referential form (Kehler et al., 2008; Fukumura & van Gompel, 2010; Rohde & Kehler, 2014). All of these studies also used a story-continuation methodology. For example, Fukumura and van Gompel (2010) examined verbs like *scared* or *feared*, which denote emotional states. They asked participants to generate continuations prompts such as *Gary scared Anna because…*, or *Anna feared Gary because…..* They asked whether participants were more likely to continue with mention of the character in the stimulus role (Gary), which is generally considered the more likely cause of the event. They found that indeed, the stimulus was mentioned more often, supporting the conclusion that thematic roles influenced preferences about **who** would be referred to. That is, they affect the referential predictability of the character. However, in none of these experiments were pronouns produced more frequently for either stimulus or experiencer roles. Instead, participants followed the first-mentioned/subject bias, preferring pronouns when they mentioned the first character (e.g. *Gary*), and names when they mentioned the second (e.g. *Anna*). The lack of a thematic role effect has led these authors to argue that predictability has no effect on reference production. Instead, they suggest that

---

[1] Note that they report their data to answer the question "Given that a pronoun is produced, what is it more likely to refer to". They do not provide data for the rate of pronoun use overall for each thematic role/grammatical role category, which is the question we ask here.

topicality is the sole determinant of pronoun selection (Kehler & Rohde, 2013; Fukumura & van

Gompel, 2010).

However, there are several reasons to reconsider the question of whether thematic roles

affect referential form, and how both are related to predictability. First, predictability has

widespread effects on other aspects of language production, such as acoustic form (e.g.,

Lieberman, 1963). Second, there are numerous differences between the Arnold (2001)

goal/source study and the studies that found no effects on referential form. This raises questions

about whether the Arnold (2001) finding was spurious, or whether the detection of thematic role

effects is dependent on other differences between the studies. Moreover, all of the previous

studies used a story-continuation methodology, raising questions about whether the effect is

robust across other methodologies. This paper therefore examines this question systematically,

using both picture-description and story-continuation tasks.

Why might predictability matter?

Predictability plays a central role in current theories of both language production and

comprehension. The probability of words and phrases is central to probabilistic models of

language processing (e.g., Jurafsky, 1996; Seidenberg & MacDonald, 1999; Hale, 2001), and

there is extensive evidence that predictability affects language comprehension (e.g., Altmann &

Kamide, 1999, van Berkum, Brown, Zwitserlood, Kooijman, Hagoort, 2005; Staub & Clifton,

2006; Levy, 2008).

Even more relevant to our question, it is well established that predictability affects

language production, by modulating word and vowel duration and consonant cluster reduction.

Over time this results in the shortening of words that are frequent (Zipf, 1936) or tend to be

produced in probable contexts Piantadosi, Tily, and Gibson (2011). The token-by-token

pronunciation of words also varies by context. Both function and content words tend to be produced in a more reduced manner when they are predictable in context, as measured by vowel duration (Bell, Jurafsky, Fosler-Lussier, Girand, Gregory, & Gildea, 2003), word duration and final consonant deletion (Gregory, Raymond, Bell, Fosler-Lussier, & Jurafsky 1999; Jurafsky, Bell, Brenier, Gregory, Girand, & Jurafsky, 2009).

The relationship between the probability of a word and how it is pronounced has been formalized in the Probabilistic Reduction Hypothesis (Bell et al., 2009; Gregory et al., 1999; Jurafsky et al., 1998), which states that the probability of a word in context predicts its degree of reduction. Information-theoretic models more broadly suggest that linguistic form is related to the information expressed by the linguistic item (Aylett & Turk, 2004; Levy & Jaeger, 2007; Jaeger, 2006). According to this view, 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 accounts for multiple linguistic phenomena: overall speech rate (Aylett & Turk, 2004), phonemic production (Son & van Santen, 2005); the use of optional function words (Jaeger, 2006; Levy & Jaeger, 2007); and contraction production (*you are* vs. *you're*, (Frank & Jaeger, 2008). When extended to referential expressions, 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. Indeed, word duration has also been shown to be sensitive to thematic role (Kaiser et al, 2011; Zerkle, Rosa, & Arnold, under review).

When we consider the selection of pronouns vs. other forms, the relevant level of predictability is the referent itself – that is, how likely is it that a particular entity will be mentioned? This contrasts with most work on word pronunciation, which has mostly focused on

the probability of words (but see Gahl & Garnsey, 2004). Thus, the question is whether the predictability of **reference to a particular entity** affects the likelihood of using a pronoun to refer to that entity (see also Kehler & Rohde, 2013).

There are many factors that might influence the predictability of a referent in the upcoming discourse. For example, the salience of an event, such as a person falling down, may increase the expectation that the event will be mentioned. Here we examine predictability as it relates to the thematic roles, which influence predictions about next-mention likelihood.

Methodological issues

Despite the support for predictability effects in language, there is mixed evidence in the literature for thematic role effects on reference production. We consider several methodological issues that may explain differences amongst existing studies.

Verb type. Most prior studies have investigated the effect of thematic role on pronoun use by examining verbs for which one argument is assumed to be implicitly more causal (Fukumura & van Gompel, 2010; Rohde & Kehler, 2014; Kehler et al. 2008 Exp. 3; Kehler & Rohde, 2013). By contrast, the one study that did find an effect of thematic role used transfer verbs, while controlling for grammatical role (Arnold, 2001). Although Kehler et al. (2008) and Rohde (2008) have also examined transfer verbs, they did not include a control for grammatical role. These verbs differ in many ways. Many of the implicit causality studies used experiencer/stimulus verbs, which denote psychological states (such as *admired* and *blamed*), and are less imageable than discrete actions (*handed, gave*). The stimulus role is considered the implicit cause of these events, but the experiencer may be accessible due to the focus on that person's mental state.

Controlling for grammatical role effects. It is well established that speakers tend to use pronouns when referring to the grammatical subject. Transfer verbs provide a good test case for predictability effects, because they allow for the effects of thematic roles to be separated from grammatical role. Some transfer verbs place the goal in subject position (get, receive, take) while others place the goal in the nonsubject position (give, hand, send). For example, in "*Bob handed the threatening note to Sue*", the grammatical subject is the source, and the nonsubject (i.e., object of preposition) is the goal. In a sentence like "*Larry got the romantic note from Ellen*", the subject is the goal and the nonsubject (object of preposition) is the source.

Given the known subjecthood bias in pronoun production, we would expect pronouns to be used for both Bob and Larry. If the predictability of goals also affects pronoun use, we would expect relatively more pronouns for Larry than Bob, and for Sue than Ellen, on top of the subject effect. This pattern was observed by Arnold (2001), except that the goal/source difference only emerged for the nonsubject references, for which speakers used around 18% pronouns for nonsubject goals, but only 7% pronouns for nonsubject sources. By contrast, pronoun use was at ceiling for reference to the subject. This finding is consistent with both a strong role for grammatical role, and a contributing effect of thematic roles.

Conversely, Rohde (2008, Exp. VII) compared goal and source thematic roles across different grammatical roles. She examined pronoun use following transfer verb prompts such as *John handed a book to Mary…*. However, in her stimuli the goal was always the nonsubject, and the source was always the subject. Participants were more likely to provide goal continuations than source continuations, but they used more pronouns for the subject than object. Rohde (see also Kehler et al., 2008) focuses on the discrepancy between the continuation bias and the rate of pronoun use. That is, if speakers tend to talk about Mary more than John, but use pronouns for

John more than Mary, it suggests that pronoun use is influenced by more than pure predictability estimations. However, this design does not allow for a test of the hypothesis that thematic role has a partial effect on reference form choice, when grammatical role is held constant.

Task demands in the story continuation paradigm. Without exception, the effect of thematic roles on pronoun use has been examined with a story continuation paradigm in every study published (to our knowledge). In standard story continuation studies, participants read a probe sentence (or sentences), and generate a continuation to the story (Arnold, 2001; Fukumura & van Gompel, 2010; Kaiser et al., 2011; Kehler et al., 2008; Rohde & Kehler, 2014; Stevenson et al., 2004). The advantage to this paradigm is that it allows tight control of the linguistic context, and can test both 1) which character is mentioned more (as a measure of predictability), and 2) what referential form is used.

Yet there is also reason to believe that the story continuation paradigm has task demands that differ from normal language production, and these demands may interfere with the researcher's ability to detect thematic role effects on reference form choice production (see Arnold, 2013, for a critique of this paradigm). Critically, in continuation tasks participants don't know what the second event is until they invent it, which means they cannot know the coherence relation between the utterances. Nevertheless, formulation of the response is likely to occur in parallel with message generation, drawing on information that is already in the context (e.g., subjecthood), but not coherence-driven semantic information. By contrast, in natural language people often know the gist of their upcoming utterances, even before linguistic formulation occurs, so the coherence relation may also be pre-activated.

Second, sentence continuation tasks are typically more de-contextualized than natural language. If conceptual integration between utterances is important for predictability effects, we

may also expect to see stronger effects of thematic roles within a richer discourse context. All

the studies that found no effect of thematic role on reference form have used single-sentence

contexts. By contrast, Arnold (2001) used a three-sentence context that ended with the critical

transfer verb probe, e.g.: *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. It may be that

having a stronger discourse context provides the conceptual support for generating predictions

about who will be mentioned next earlier in the response process.

Current study goals and methodological approach

This paper reports the results of a systematic investigation of whether thematic roles

influence reference production, and if so, whether this effect can be attributed to the

predictability of certain referential continuations. We examined transfer verbs, which have been

shown to influence the use of pronouns (Arnold, 2001). Our primary goal was to test whether

this effect was real, given widespread claims that pronoun use is unaffected by either the

thematic roles or the predictability of potential referents (Kehler et al., 2008; Kehler & Rohde,

2013; Fukumura & van Gompel, 2010). Our secondary goal was to consider possible

mechanisms by which predictability might affect reference production.

Given the methodological concerns about the sentence continuation paradigm described

above, we created a novel event-retelling paradigm, and compared it with sentence-continuation

studies. In the event-retelling paradigm, participants were asked to describe a series of picture

pairs, which together told the story of Clue-like murder mystery. The participant was given the

role of tabloid photographer, and asked to describe "their pictures" to a detective who was trying

to solve a murder mystery. We asked them to preview all the pictures for the entire story, before

beginning the picture-description task. This made our task consistent with real-world scenarios in

which people know the content of their message before they produce it, as well as its role in the

story. This task was designed to be engaging and interesting for task participants, and encourage

them to develop a richer discourse representation. The story featured three main male characters

(Sir Barnes, the chauffeur, and the butler), and three female (Lady Mannerly, the chef, and the

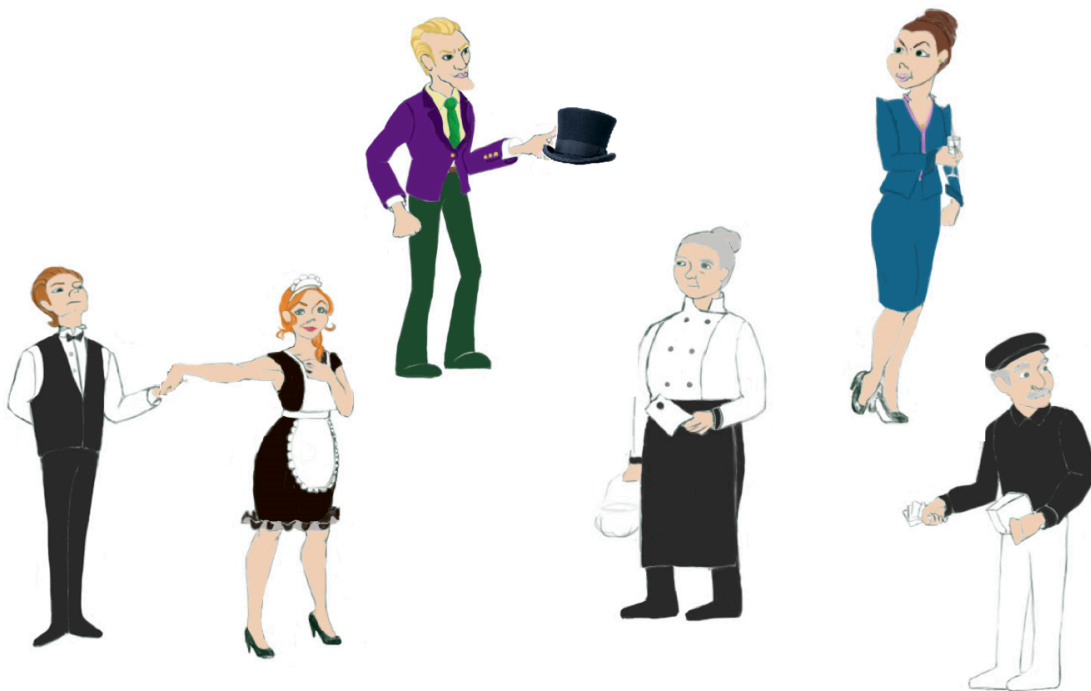maid; see Figure 1). The characters' behaviors and actions were consistent with their real-world

roles.

Figure 1. Characters in the event-retelling paradigm (from left-to-right: The butler and the maid,

Sir Barnes, the chef, Lady Mannerly, the chauffeur).

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). This paradigm had several advantages. First, it utilized a typical trial-by-trial experimental structure, while still retaining the coherence of a naturalistic storytelling situation. Second, this paradigm allowed us to manipulate the linguistic context. In each trial, two pictures were presented. The detective (an experimenter) described picture 1, allowing us to control the linguistic form of the context sentence. The subject described picture 2. Third, we were able to control the content of the participant's responses through the pictures, such that the continuation mentioned either the goal or source character. Fourth, our paradigm allowed us measure the latency between the picture onset and the participant's response, which supported inferences about the underlying mechanism.

A potential limitation to our story-telling paradigm is that the items are not entirely independent of one another, since together they tell a story. However, the benefits from this property were judged to outweigh the non-independence of items. Recent work has demonstrated that even with unrelated stimuli, subjects pick up on experiment-wide patterns and often change their behavior over the course of the experiment (Fine, Jaeger, Farmer, & Qian, 2013). We took care to consider this effect by including trial order as a fixed effect in the models.

Our study focuses on the speaker's use of pronouns, compared with the use of more explicit names or titles. Pronouns represent one of the most reduced expressions available in English. However, speakers also have other choices. For example, they can choose to not begin a new sentence, and instead use a VP-coordination structure, e.g. *and pointed to a cake recipe inside*. This structure avoids using a name or description, and plays a similar pragmatic function to the use of pronouns, except it is usually limited to cases where the subject is the same as in the previous sentence. However, this structure was rarely chosen in the current set of studies (12 trials in Exp. 1, 0 in Exp. 2, and 1 in Exp. 3), and therefore is excluded from analysis.

Our experiment design also took care to avoid ceiling or floor effects in the data. The use of reduced forms is influenced by much more than just thematic and grammatical roles, and can be heavily influenced by details of the task and instructions. Moreover, participants may individually adopt modes of speaking in which they use either only pronouns or only names. If so, it compromises our ability to detect effects of the linguistic context, in that there is no variation in responses. To avoid these problems, we excluded participants who performed at floor or ceiling (for a similar convention, see Filmer, Mattingly, Dux, 2015; Buschkuehl, Hernandez-Garcia, Jaeggi, Bernard, & Jonides, 2014), by only including data from participants who at least two pronouns and at least two names in the critical items. 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 such participants might mask the thematic role effect. At the same time, we verified that our exclusion criteria did not create a spurious effect. All analyses were repeated, using the entire set of usable subjects (i.e., those that were native speakers), and reported in footnotes for each experiment.

We also sought to avoid ceiling or floor effects by manipulating gender ambiguity in our utterances. It is well known that speakers use pronouns more often when there is only a single referent in the context that matches the pronoun's gender (e.g., *Bill called Sue and she...*) than when the pronoun would be ambiguous (*Amy called Sue and she…*; Arnold & Griffin, 2007; Fukumura, Hyönä, & Scholfield, 2010). Depending on task-specific biases, this may lead to either ubiquitous use of pronouns for different-gender contexts, or ubiquitous use of names for same-gender contexts. Since we did not know ahead of time how participants would respond to our task, we included half same-gender and half different-gender contexts.

Experiment 1 describes our in-person event-retelling experiment. In Experiments 2 and 3, we compare the results on Experiment 1 to two story-continuation experiments, to assess the role of task demands on thematic role effects in reference production. In order to examine the relationship between thematic roles and predictability in our materials, we also conducted two rating studies, which are described before Experiment 1.

General study design

The stimuli for this study were designed with verbs that describe transfer events, which can include two human characters in the roles of goal and source. Examples 3 and 4 depict two sample items. The underlined character is the one who is pictured in the second image in Experiment 1, and thus the character who is continued. Half the critical items (N=12) continued with the goal, and half (N=12) continued with the source (between-items manipulation). Within each condition, half the continuations occurred in a context where the referent was in subject position, and half in nonsubject position (within-items manipulation).

(3) Goal continuation

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

   b. Nonsubject position: *Lady Mannerly gave a backrub to Sir Barnes*.

(4) Source continuation

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

   b. Nonsubject position: *The maid took a cookbook from the chef*.

In sum, there were four conditions in the experiment: (1) Goal, subject reference; (2) Goal, nonsubject reference; (3) Source, subject reference; and (4) Source, nonsubject reference.

Figure 2. Sample trial from Experiment 1

Our interest in transfer verbs stems from evidence that the goal argument is considered more likely to be mentioned again in a subsequent event (e.g., Stevenson et al., 1994). This expectation is strongest when the following utterance describes the next event in a narrative. We therefore designed the study within a narrative context, so all items represented a sequence of events. This supported the activation of coherence relations that focus on the following event or consequence.

In the event-retelling paradigm (Experiment 1) participants heard the prompt sentences and saw the pictures, as in Figure 2. In the rating studies they read the sentences and saw the pictures. In the sentence completion paradigms (Experiments 2 and 3) participants read the prompt sentences, without pictures.

The 24 critical stimuli were arranged following a Latin Square design, where each participant was exposed to each item in only one condition, but saw all conditions across

different items. This design also encouraged variation in the pattern of reference across the items, which helped discourage participants from falling into a repetitive pattern of responses. The 29 filler items helped develop the storyline and added variety to the kinds of sentence structures encountered.

General analytic approach

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. Generalized linear mixed-effects models were used to account for the dependencies in the repeated measures. We used a logistic regression (SAS proc glimmix) for analyses of dichotomous outcomes, and a mixed-effects linear regression (SAS proc mixed) for analyses of continuous outcomes.

In order to avoid over-fitting our model, we first built a model that included only the critical predictors (subject/nonsubject, 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. Goal/source, gender, and subject/nonsubject were centered by coding them as 1/0 and grand-mean centering.

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, we then added the interactions Subject*Goal, Goal*Gender, Subject*Gender, and Subject*Goal*Gender in a second model. If any interactions approached significance (with $t > 1.5$), they were retained for the final model; all other interactions were eliminated. Here we report just the final model, although the same pattern of significance obtains whether all the interactions are included or not.

All models included random intercepts for participants. Random slopes for participants by subject/nonsubject and goal/source were included initially, but were deleted if the slope was estimated to be zero by the model. Since item/order was included as a fixed effect in all critical models it was not utilized as a random intercept in any model.

Experiment 1: In-person study

Method

Participants

32 undergraduates completed the task for class credit. 9 participants were excluded for using fewer than two pronouns and two were excluded for being non-native English speakers. This left 21 participants in the analysis, 11 on List A and 10 on List B.

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 an experimenter playing the role of detective. Participants then provided their own description of the second picture in the pair. The stimuli were divided across two lists, such that all participants saw the same pictures, but heard one of two versions of the critical prompt sentences. Experiment 1 stimulus pictures can be found on the supplementary materials website[2].

Rating studies

---

[2] jaapstimuli.web.unc.edu

Two rating studies were conducted using the stimuli for Experiment 1, to examine the relationship between our stimuli and different measures of predictability. Twenty participants completed each of the rating studies.

Next-Mention Biases. A rating study determined participants' next-mention biases. 20 participants viewed the first sentence and picture of each of the 53 stimuli and filler pairs, and selected which character they thought would be more likely to be talked about next[3]. The probability of choosing the goal was 71%, supporting the predictability of goals. The probability of choosing the subject was 54%, suggesting that subjects were not considered more predictable than nonsubjects. We tested these effects in a logistic regression where the dependent measure was the choice of the "target" (i.e., what would be the target in exp. 1). The model included predictors for both 1) subjecthood and 2) thematic role, and random slopes for each predictor by participant. Target choice was more likely when it was a goal than a source (t(19)=4.91, p<.0001), but there was no effect of subjecthood (p > .1).

Relatedness and Predictability. Another rating study examined how related and predictable the participants thought the events in the pairs of pictures were. A different set of 20 participants viewed the 53 sentences and pairs of pictures. Using a 7-point scale, they rated the pairs for: (1) how related the second event was to the first and (2) how predictable it was based on the first. Scores were converted to Z-scores based on the participant mean and standard deviation, in order to account for differences in how participants used the rating scale. These

---

[3] A research assistant gave these instructions orally to participants: "Now you're going to see some of those pictures you took, along with descriptions of them. Your job is to read the description and look at the picture and decide who you think, based on that picture, would be most likely to be talked about next. So, you'll be deciding between two characters, and you'll be choosing which one you think would be talked about next, based on the picture and description of that picture that you see. Remember to both read the descriptions and look at the pictures. Do you have any questions?"

scores were examined in a mixed-effects linear regression with two predictors: 1) subjecthood, and 2) thematic role, and random slopes for each predictor by participant.

The items in which the second picture featured the goal were judged to be more related (mean rating = 5.47) than the items in which the second picture that featured the source (mean rating = 4.81), as confirmed by a main effect of thematic role ($\beta = 0.34$ (.09), t = 3.65, p = .002). There was no main effect of subjecthood (p > .8), and no interaction when it was added to the model. There was a numerical trend for goal conditions to be judged to be more predictable (mean rating = 3.825) compared to source conditions (mean rating = 3.57). However, neither predictor reached significance in the model (t's < 1.1, p's > .2). Note that this metric reflects the predictability of the second event overall, and not specifically the predictability of reference to the target character, which was measured in the first rating study.

In sum, the goal character is considered more likely to be mentioned than the source character –that is, more predictable. The goal-continuation events are also perceived as more related to the context than the source-continuations, but the events themselves are equally predictable.

Event Retelling 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 background slideshow. 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. They then completed

a sample item with the experimenter. The experimenter explained that the detective, who would

arrive shortly, would describe the first picture in the pair. The detective was presented as a part

of the role-playing game, and not a naïve participant. After that, the participant should say what

happened next, using the second picture as a guide.
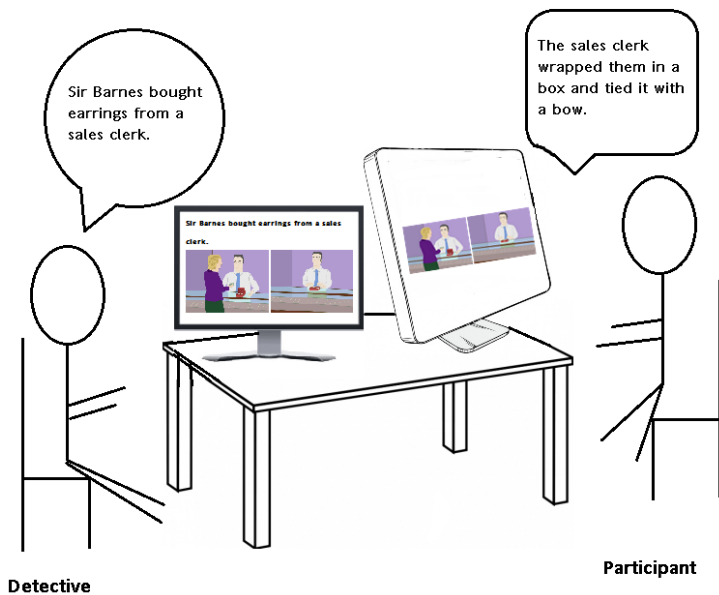
Figure 3. Experimental set-up from Experiment 1

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. The two computers were placed back-to-back, and 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, and the subject 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 3. 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.

Analysis

Post-experiment questionnaire

A post-experiment questionnaire confirmed that most participants (18/21) were familiar with the Clue game, which is likely to have supported their development of rich mental models. This questionnaire also motivated participants' engagement with the task. 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, which they were asked in the post-experiment questionnaire. All participants correctly identified Sir Barnes as having been the character that was killed, and the motive as having something to do with Sir Barnes' affair with the maid. Third, we asked participants what they thought the goal of the experiment was. Participants reported several possibilities but none identified the specific linguistic manipulations.

Response coding

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 codings for Experiments 2, which had been run first, no double-coding was performed for this experiment.

72 trials were excluded from the final analysis, leaving 432 trials in the analysis. 53 trials were excluded because the target character was not mentioned as the subject of the response, one was excluded for using 'who' as the Subject, twelve for using a VP-coordination structure for the response (e.g., *and then wrapped them up in a gift box),* and six were excluded due to mechanical issues (two pictures were advanced instead of one; the picture was advanced too soon, etc.).

Audio data coding

The audio data were analyzed with Praat to measure latency to begin speaking, defined as the end of the detective's speech to the beginning of the subject's response. These time points were first coded both by two undergraduate research assistants, and the first author checked those items that differed more than 10% (n=100, 23% of total),[4] and used the measurement of the first RA for those that did not. The entire dataset was also coded by a third RA, to double-check the dataset for reliability. The latency measurement was log transformed, and we excluded the one outlier that exceeded the mean plus 3 standard deviations of the entire dataset.

Results

Semantic and grammatical role effects on pronoun production

The critical question was whether participants would use more pronouns to refer to the goal of the prior sentence as compared to the source. Indeed, pronouns were more common for reference to prior goals as compared to prior sources, as can be seen in Figure 4 and Tables 1 and 2. As was expected, they also used more pronouns when referring to Subjects of the prior sentence as compared to nonsubjects. There was also a marginally significant interaction

---

[4] 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 disfluent fillers (*um* was to be included in the latency period, rather than as the onset of the speech).

between subjecthood and thematic role. Contrasts in the model revealed that the thematic role effect was robust for the subject condition (t = 3.79, p = .0002), but not the nonsubject condition (t = 1.21, p = .225). There was also a general trend to use fewer pronouns toward the end of the experiment (Order effect), but no effect of same- vs. different-gender.
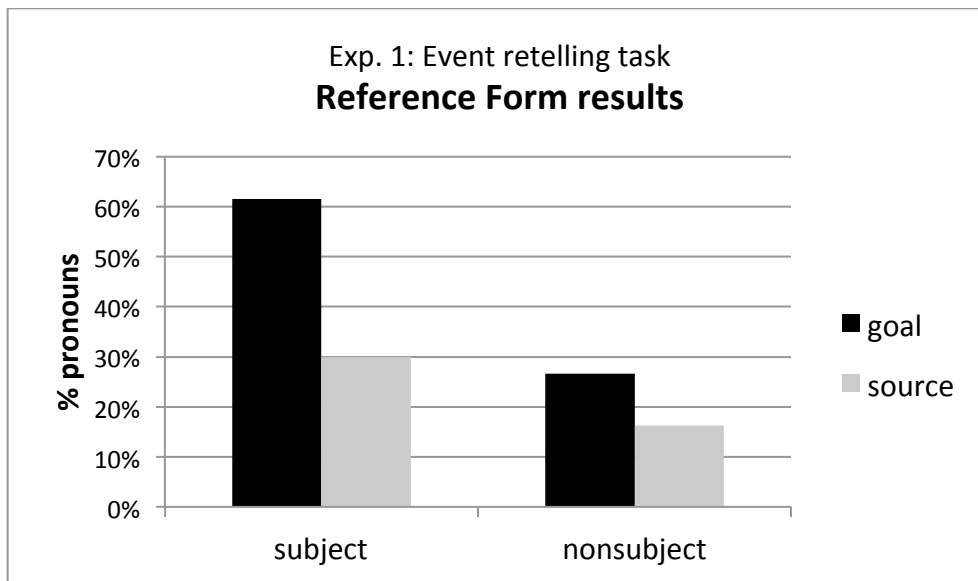

Figure 4. Percentage of pronouns used by semantic and grammatical roles in prior sentence in Experiment 1

Table 1. Experiment 1 results: Rate of pronoun production by condition

		goal	source
different-gender	subject	64%	37%
	nonsubject	31%	18%
same-gender	subject	59%	22%
	nonsubject	23%	15%
OVERALL		43%	23%

Table 2. Experiment 1 Results: Pronoun rate model (including interactionswith t > 1.5). This model has random slopes for both subjecthood and thematic role by participant.[5]

Variable	Estimate (Error)	t-value	p-value
Subject vs. Nonsubject	1.41 (0.32)	4.41	0.0003
Goal vs. Source	0.93 (0.30)	3.04	0.006
Order	-0.02 (0.008)	-2.60	0.01
Same gender vs. Different gender	-0.39 (0.24)	-1.62	0.11
Subjecthood x Goal	0.94 (0.50)	1.89	0.06

Latency effects

If goal continuations are easier to plan and produce, we would expect to see an effect on participants' latencies to begin speaking. Indeed, utterance initiation latencies were shorter when the participant was referring to a goal as opposed to a source. In addition, the same-gender items elicited longer latencies than the different-gender items. There was a marginal interaction between Gender and Thematic role, but contrasts revealed a significant goal effect for both same- and different-gender items (Same: t = -2.13, p = .03; Different: t = -4.88, p < .001). There was no effect of referring to the subject versus the nonsubject, and no interaction between the two. Latency data can be seen in Figure 5 and Table 3.

---

[5] We also analyzed the full set of data from the 30 participants who were native speakers. This includes 6 participants who used only names/descriptions, and 3 who used only 1 pronoun. The pattern of significance was identical to that shown in table 2, demonstrating that our system of excluding participants did not create an effect where there wasn't one.
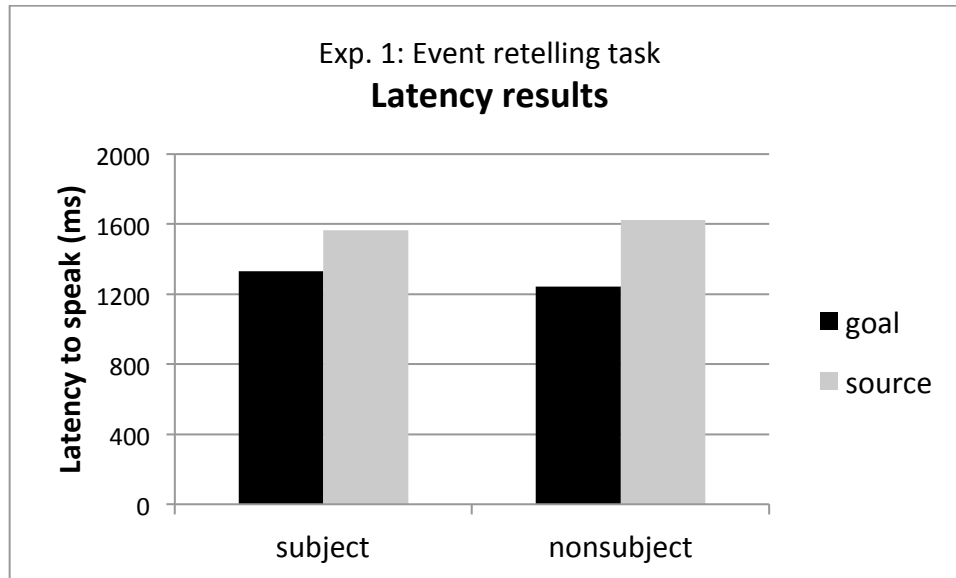
Figure 5. Latency to begin speaking by grammatical and semantic roles in prior sentence in

Experiment 1

Table 3. Experiment 1 results: Latency model (including interactions significant at t>=1.5).
This model has random slopes for both subjecthood and thematic role by participant.

Variable	Estimate (Error)	t-value	p-value
Subject vs. Nonsubject	0.01 (0.02)	0.29	0.77
Goal vs. Source	-0.1 (0.02)	-4.81	<.0001
Order	0.001 (0.0007)	1.59	0.11
Same gender vs. Different gender	0.04 (0.02)	1.97	0.0497
Same/diff gender x Goal	0.07 (0.04)	1.87	0.062

Given that goal continuations yielded faster response latencies, and were also more likely

to involve pronouns, we asked whether latency itself was a predictor of reference form, and

possibly the mediator of the goal effect. However, when we added latency to the reference form

model, we found that it was not significant, while the other predictors revealed the same pattern

as reported in table 2. Even though latency was moderately correlated with thematic role (r = -

.21, p < .001), it was not a significant predictor even in a model by itself.

Discussion

Our first question was whether thematic roles influence the production of reduced referential expressions. We found that indeed they do: speakers used pronouns for goal referents more than source referents. This effect was strongest for the grammatical subject. This thematic role bias sat on top of the expected subject bias, where pronouns were more frequent for reference to subjects than nonsubjects.

Our second question was whether these effects were related to the predictability of goals. Our rating study revealed that the goal character in our stimuli was considered more likely to be mentioned again than the source character, confirming reports that references to goals are predictable (Arnold, 2001; Stevenson et al., 1994). This suggests that predictability underlies our thematic role effect.

If predictability does affect reference production, the next question is why. One hypothesis we considered was that both pronouns are more likely in situations where production planning is easier. If it is easy for the speaker to activate representations of the events to be described, it may encourage both earlier speech and greater connection between the utterances. Consistent with this, we found that goal utterances were easier to plan than source utterances, in that they required shorter latencies to respond. However, when latency was added to the model for reference form choice, it did not independently predict reference form choice. This may indicate that utterance planning is not directly related to reference form choice. However, our task was not designed to allow a fine-grained measure of planning time. Participants had previewed the pictures, and further were able to begin planning their response during the detective's description of the first picture, such that the latency was only a rough measure of

planning difficulty. Thus, further work is needed to precisely specify the relationship between latency and reference production.

Our observation of a thematic role effect stands in contrast to several published studies. This raises questions about what accounts for the differences. As described above, our event-retelling task differed from previous paradigms on two important dimensions. First, participants were not required to invent a continuation, and instead described a picture that they were already familiar with. Second, our items together told a story, increasing the ability to represent the discourse context. We examined each of these differences in Experiments 2 and 3.

Experiment 2

Experiment 2 was conducted using the same linguistic material as Experiment 1 but with the written story-continuation paradigm, to determine whether the same effects could be found with the standard methodology. We used a variant of the method used by Fukumura and van Gompel (2010, experiment 1), and told participants that they had to begin their continuation with the underlined character.[6]

Method

Participants

20 participants were included in the analysis; an additional 26 completed the task but were replaced. Of these 46 participants, 10 were undergraduates who were reimbursed with course credit and 36 were recruited via Amazon Mechanical Turk and received monetary compensation. The Amazon Mechanical Turk participants needed to be native English speakers with a HIT approval rate greater than or equal to 98%, with at least 5000 approved HITs. The

---

[6] Another experiment, reported in Rosa (2015), used the same story-continuation paradigm, but did not tell participants who to mention. The numerical patterns in this experiment were very similar to those reported here, but the frequency of continuing with the goal character was so strong that there were very few source-continuation items, which hindered our ability to assess the effect statistically.

undergraduate participants needed to be native English speakers, have normal or corrected-to-

normal vision, and couldn't have participated in a similar experiment in the lab. 8 participants

were excluded for using fewer than 2 proper names, 17 for using fewer than 2 pronouns (one of

these also had meaningless responses), and 1 because we had collected enough data for even

numbers on each list.

Materials, Design, and Procedure

The experimental trials consisted of the first in each pair of sentences from the storyline,

which included 24 critical items and 29 fillers. 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, which they typed in the

box. Although the stimuli sentences were identical to the context sentences from Experiment 1,

this task did not include any instructions to place the task within a murder-mystery narrative.

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. An example item from Experiment 2

is shown in Figure 6:

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

Figure 6. Sample trial from Experiment 2. Underlining indicates which character should be

continued.

Analysis

Response coding

 We coded the grammatical subject of the first clause of each response, where a clause could be either a main or subordinate clause. Responses were coded for both a) choice of referring expression (pronoun or proper name) and b) role of referent in the prior sentence (subject or nonsubject 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, or they referred to someone's possession or body part as the subject (e.g., *Sir Barnes' back was sore*).

 Data were double coded by the first author and two undergraduate research assistants who were blind to the original coding. Of the 480 items, six mismatched. The first author determined one of these to be about the wrong referent, and excluded the other five because they were ambiguous. Following the criteria mentioned above, fifty-five items were excluded from the final analysis, leaving 425 items (223 goal items and 202 source items; 223 subject items and 202 nonsubject items).

 The chef, who was pictured as female in Experiment 1, was interpreted as male by most of the participants. Likewise, the sales clerk was intended to be male but most participants interpreted him as a female. We changed the coding of gender for the items with the chef or the sales clerk to reflect this.

Coherence relations coding

 The relationships between the prompts and the continuations given were also examined, given previous claims that coherence relation affects the predictability of thematic roles. Using

Rohde's coding schema,[7] two undergraduate research assistants independently coded each continuation. The seven categories they used were elaboration, explanation, occasion, parallel, result, violated expectations, and background. After coding all the items, the RAs then compared their ratings. On 147 of the 425 total they had coded items differently, so these items were discussed until they had reached an agreement about the appropriate coding.

Two of the continuation types (result, occasion) describe events that occur as a result of another event or after it temporally, and thus are more consistent with goal as opposed to source continuations (Rohde, Kehler, & Elman, 2007). We therefore collapsed the continuation codings into two groups: a) Occasion/Result and b) Other. Of the goal continuations, 120 were coded as Occasion or Result, and 103 were coded as a different continuation type. Of the source continuations, 81 were coded as Occasion or Result and 121 were coded as a different type.

Statistical modeling

The data were analyzed following the general analytic approach outlined above, except that coherence continuation was added as a control variable. We also examined interactions between coherence continuation and the critical goal/source variable.

Results

Semantic and grammatical role effects on pronoun production

Similar to Experiment 1, participants used more pronouns when referring to goals (63%) of the prior sentence as opposed to sources (51%). As expected, participants used more pronouns when referring to Subjects (76%) of the prior sentence as compared to nonsubjects (36%). We found a main effect of gender context, such that pronouns were more frequent in the different-gender items (62%) than the same-gender items (53%). There was a significant effect of item

---

[7] We are very grateful to Hannah Rohde for sharing her coding schema with us for this project.

order, but in the opposite direction from Experiment 1: for Experiment 2 pronoun use increased

slightly for later items. Additional models confirmed there were no interactions amongst the

subject, goal, and gender variables. Model results are presented in Table 5.

Table 4. Experiment 2 results: Rate of pronoun production by condition

		goal	source
different-gender	subject	83%	78%
	nonsubject	55%	33%
same-gender	subject	69%	69%
	nonsubject	18%	19%
OVERALL		63%	51%

Table 5. Model of the production of pronouns in Experiment 2. The model had a random slope for subjecthood by participant. The random slope for goal was estimated to be zero.[8]

Variable	Estimate (Error)	t-value	p-value
Subject vs. Nonsubject	2.82 (0.34)	8.35	<.0001
Goal vs. Source	0.75 (0.29)	2.62	0.009
Order	0.03 (0.01)	2.6	0.01
Same gender vs. Different gender	-1.63 (0.33)	-4.91	<.0001

To examine the effect of coherence relation, we added it to the model, along with

interactions between Occasion/Result continuation and both thematic role and Subjecthood (see

Figure 8). In addition to the main effects shown in Table 6, this model revealed a marginal

interaction between thematic role and coherence relation. There was no main effect of coherence

relation, nor interaction between coherence relation and subjecthood (p's > 0.1). Contrasts

revealed the marginal thematic role x coherence relation interaction was due to a significant

difference in pronoun use for Occasion/Result continuations between goal and source items (β =

1.22 (SE = .40), t=3.06, p = .002). No difference in pronoun use was found between goal and

source items for Other continuations (β = .86(0.55), t= 1.56, p = 0.12) (see Figure 8).

---

[8] We also analyzed the full set of 45 participants (not including one who produced meaningless responses). The pattern of significance was identical.

Table 6. Experiment 2 results: Pronoun production model, including coherence relations. The model had a random slope for subjecthood by participants. The random slope for goal was estimated to be zero by the model.

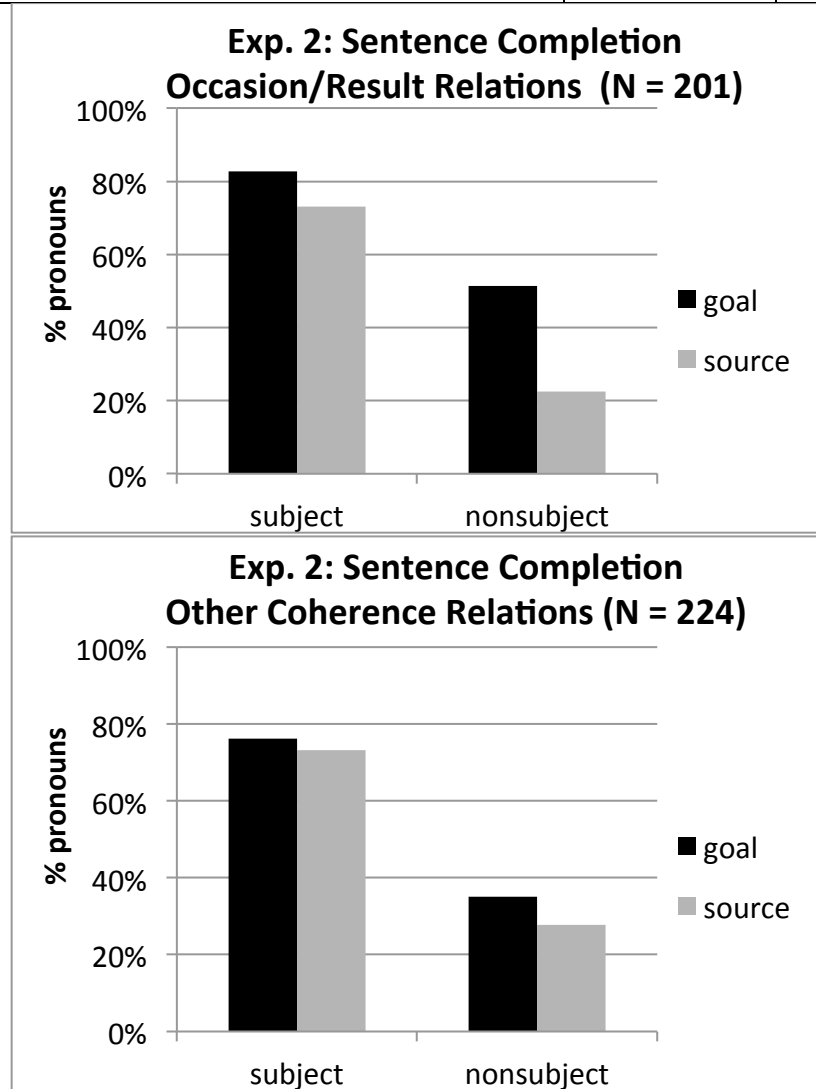| Variable	Estimate (Error)	t-value	p-value
Subject vs. Nonsubject	2.94 (0.35)	8.41	<.0001
Goal vs. Source	0.7 (0.29)	2.41	0.02
Order	0.03 (0.01)	2.44	0.01
Same gender vs. Different gender	-1.6 (0.33)	-4.8	<.0001
Occasion/Result vs. Other relations	0.42 (0.32)	1.34	0.18
Goal x Coherence relation	1.08 (0.57)	1.88	0.06
Subjecthood x Coherence Relation	0.47 (0.58)	0.82	0.41

Figure 8. Experiment 2 results: Proportion of pronouns used by semantic roles and grammatical function in prior sentence, split by coherence relation.

Discussion

Experiment 2 found the predicted thematic role effect. Participants produced more pronouns when referring to prior-goals than prior-sources. They also showed the expected effects of producing more pronouns for prior subjects as opposed to prior nonsubjects, and in the different-gender condition. Coherence relations somewhat modulated this effect: when the coherence relation was consistent with the thematic bias (i.e. a prior-goal in an Occasion/Result continuation), more pronouns were used than when it was inconsistent. While the observed gender effect did not show up in Experiment 1, it has been reported repeatedly in the literature (Arnold & Griffin, 2007).

Critically, this experiment demonstrated that the thematic role effect is robust to the experimental paradigm. Despite the concerns that we had about the story-continuation paradigm, it appears that it is possible to detect thematic role effects using this method. However, there was one way in which this experiment differed from the standard method, in that the items all referred to the same cast of six characters, and together told a story. Experiment 3 tested whether this relatedness was necessary to observe the thematic role effect on reference production.

Experiment 3

Experiment 3 was conducted to determine whether the thematic effect found in Experiments 1 & 2 was dependent upon the experimental items being related to one another in the story context. 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, freeing up

mental resources to utilize predictability information. Experiment 3 was identical to Experiment 2, except it eliminated the repeated mention of people and items.

Method

Participants

57 participants completed the task on Amazon Mechanical Turk, all for a monetary reward. The same inclusion criteria were used as for the M-Turk participants in Experiment 2. 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 an earlier version of the experiment, three were excluded for providing meaningless continuations, 25 were excluded for using fewer than two pronouns, and four were excluded for using fewer than two names.

Materials, Design, and Procedure

The design and procedure was identical to Experiment 2. The only difference was that the names/occupations and objects were changed, such that none were repeated across items. Names were replaced by another common name of the same gender (selected from a list of the most popular male and female names in 1958), and occupations were replaced by another 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 mentioned in the original items were replaced with other common objects such that none were repeated. An example is *Michael received a painting of the two of them from Mary*.

Analysis

Response Coding

The inclusion criteria were identical to those of Experiment 2. Given the very high consistency of ratings between the original coder and the re-coders for Experiment 2, no double coding was performed for this experiment. Fifty-nine items were excluded from the final analysis, leaving 421 items (210 in the subject condition and 211 nonsubject condition; 222 in the goal condition and 199 in the source condition).

The coding procedure was identical to Experiment 2, except that for the coherence relations coding we only coded for the binary distinction that was used for the analysis (Occasion/Result vs. other). One undergraduate research assistant coded all the data and the first author (ER) coded a sample of 20% of the data (84 items) to check for consistency. The two coders disagreed on 6 of the 83 items, or 7%. Four of these six were determined to be coded correctly by the research assistant, and the other two were determined to be coded correctly by the first author.

Results

As seen in Experiment 2, pronoun use was higher for subjects than nonsubjects, for goals than sources, and for different-gender than same-gender items (see Table 7 and Figure 9). These effects were qualified by a thematic role by subjecthood interaction, and a thematic role by gender interaction (see Table 8). Contrasts in the model suggested that the interaction between subjecthood and thematic role was due to an effect of thematic role in the nonsubject condition $t(374)=3.67, p<.0001$, and no effect in the subject condition $t(374)=0.96, p=0.33$. The thematic role by gender interaction was revealed by contrasts to be due to a significant thematic role effect in the same gender condition $t(374)=3.71, p=.0002$, and no significant thematic role effect in the

different gender condition $t(374)=0.95$ $p=.34$. As in Experiment 2, pronouns also became more
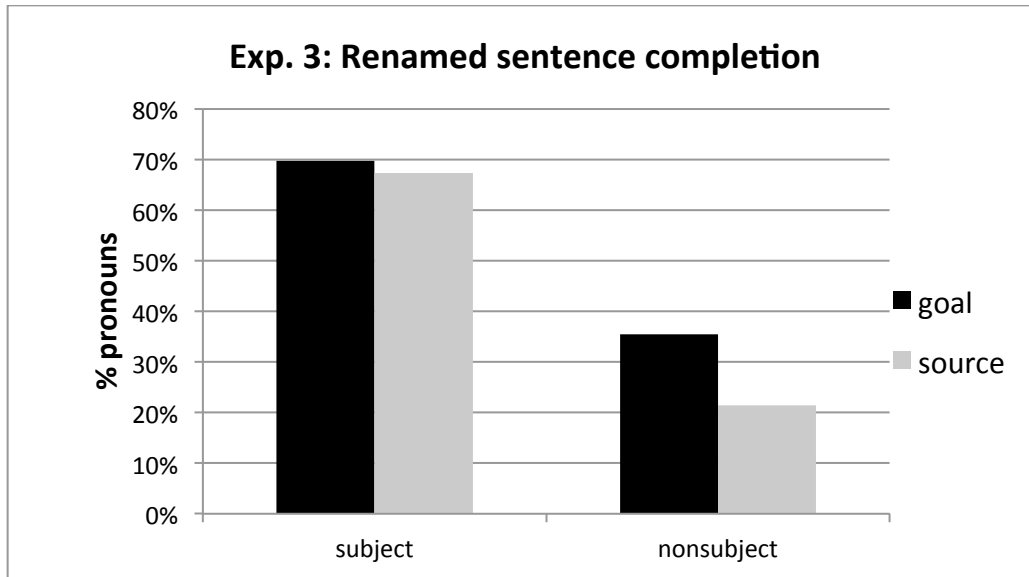
common later in the task.

Exp. 3: Renamed sentence completion

Figure 9. Proportion of pronouns in Exp. 3, by thematic and grammatical role in prior sentence.

Table 7. Experiment 3 results: Rate of pronoun production by condition

		Goal	source
different-gender	subject	71%	71%
	nonsubject	38%	33%
same-gender	subject	68%	64%
	nonsubject	33%	10%

We examined the effect of coherence relation by adding it to the model as a binary

predictor (Occasion/Result vs. Other), just as in Exp. 3, retaining only the significant interactions

from the model in Table 8. However, coherence relation had no effect, and the other effects

remained as in Table 8. Additional models confirmed that coherence relation did not interact

with either thematic role or subjecthood predictors.

Table 8. Experiment 3 results: Pronoun rate model (including interactions). This model had a random slope for subjecthood by participants; the random slope for thematic role was estimated to be zero by the model.[9]

Variable	Estimate (Error)	t-value	p-value
Subject vs. nonsubject	3.24 (0.4)	8.15	<.0001
Goal vs. Source	1.07 (0.33)	3.27	0.001
Order	0.05 (0.01)	4.58	<.0001
Same gender vs. Different gender	-0.92 (0.3)	-3.06	0.002
Subjecthood x Goal	-1.35 (0.6)	-2.25	0.025
Subjecthood x Gender	0.66 (0.59)	1.12	0.261
Goal x Gender	1.35 (0.6)	2.25	0.025
Subjecthood x Goal x Gender	-1.92 (1.19)	-1.61	0.109

Discussion

As found in Experiments 1 & 2, participants referred to prior goals with more pronouns as compared to prior sources. As expected, participants also preferred using pronouns when referring to subjects of the prior sentence as opposed to nonsubjects, and in the different-gender condition. As the experimental items were no longer related to one another, the thematic role effect could not be due to any coherent, overarching model of the events.

Experiment 3 revealed a few interactions that were not observed in Experiment 2. The goal bias was stronger for the nonsubject condition, compared with the subject condition. Notably, this interaction goes in the opposite direction from that observed for Experiment 1. Pronoun use was higher overall in Exp. 3 than Exp. 1, suggesting that the thematic role effect is strongest when participants are not at ceiling or floor within a particular task: for Exp. 1, the subject condition led to more variation, whereas in Exp. 3, the nonsubject condition led to more variation, and these were the conditions where we detected the thematic role bias.

---

[9] We also analyzed the full set of data from the 30 participants who were native speakers. The pattern of significance was identical to that shown in table 3.

In addition, the thematic role effect was stronger in the same-gender condition than the different-gender condition. These findings underscore the fact that the thematic role effects are overall relatively weak, compared to the subjecthood effect. It may also be that the lack of a coherent story here diminished the ability of participants to build a robust mental representation of the events, weakening the thematic role effect further, and possibly contributing to the lack of a coherence relation effect (unlike in Experiment 2). Nevertheless, it is notable that we again replicated the finding that pronouns are preferred for goals over sources, here in the nonsubject condition. This finding argues against the hypothesis that a story-like context is required to observe the effect.

General discussion

The findings from our experiments were strikingly consistent: In all three studies, we found that thematic roles do influence referential form. Across two paradigms and with different sets of materials, participants consistently used more pronouns to refer to the goal character than the source character. Our thematic role finding was also robust to our criteria for including participants. Additional analyses tested the effects of our primary predictors for all usable participants, and found the same pattern of results. Moreover, two subsequent experiments have used variations on this paradigm (Zerkle, Rosa, & Arnold, 2015), and replicated the preference to use reduced expressions for goal characters more than source characters.

This empirical finding firmly establishes that thematic roles can indeed influence decisions about referential expression. This finding is important, given that several researchers have claimed that thematic roles are irrelevant to reference form production (Fukumura & van Gompel, 2010; Kehler & Rohde, 2013; Kehler et al., 2008). Moreover, we found evidence

against our hypothesis that thematic role effects would only show up within rich, interactive

story-telling tasks.

We also observed the expected interaction between coherence relation and thematic role

in experiment 2 (although it was only marginal), where the goal bias was strongest in the

Occasion/Result continuations. This supports the claim that a sentence about the result or next

event encourages a focus on the goal (Stevenson et al., 1994). Importantly, this suggests that

contrary to our hypothesis, even in a story continuation paradigm, participants are able to

generate a continuation and activate the coherence relation early enough for this to affect the

selection of a referential form.

Even though our effects were robust across tasks, we still saw important differences in

reference production across the experiments. The biggest difference was between in-person

(Exp. 1) and internet-based tasks (Exps. 2 and 3). Pronouns were much less frequent in Exp. 1

(33%), and participants were excluded more often for not using enough pronouns than using too

many. By contrast, pronouns were more common in Exps. 2 (57%) and 3 (49%), where

participants tended to be excluded for using too many pronouns, and also tended to increase their

use of pronouns toward the end of the task. These participants probably wanted to finish as

quickly as possible, and realized that pronouns were faster to type. In addition, the picture-

description task in Exp. 1 carries an implicit goal to provide accurate descriptions – a goal that is

encouraged by our procedure of testing participants on the names of the characters, to make sure

they remember them. This variability across tasks is important, because thematic role effects are

more likely to emerge in the linguistic conditions that bring the participant away from floor or

ceiling. This led to different interactions across tasks, such that we saw more pronouns for goals

in Experiment 1 for subjects, in Experiment 2 for both subjects and nonsubjects, and in

Experiment 3 for nonsubjects (and same-gender items).

Our tasks also differed in the strength of the story context. The difference between the

goal and source conditions was larger in those experiments that were organized around a murder-

mystery story – namely, Experiments 1 (19%) and 2 (12%) – than the one that used disconnected

items, Experiment 3 (8%). This suggests that while a story-like context may not be essential for

thematic role effects on reference form, it may be helpful.

What explains the contrast with previous reports?

Our results stand in contrast to evidence from several studies that thematic role biases do

not affect the production of pronouns (Fukumura & van Gompel, 2010; Kehler et al., 2008).

There are several possible explanations for this. We ruled out the sentence-completion task as the

culprit, and showed that longer story contexts are not required either. Yet several differences

remain between our study and those that found no thematic role effects.

The main difference is that the current study used transfer verbs, while most other studies

have examined implicit causality effects in emotion verbs like *admire* or *impress*, which have

experiencer/stimulus roles. We speculate that the thematic role effect may be restricted to

transfer verbs. Transfer verbs differ from experiencer/stimulus verbs in multiple ways. First,

transfer verbs are telic (they have an endpoint), whereas the experiencer/stimulus verbs are atelic

(they do not have an endpoint). Telic events may be easier to conceptualize, leading to a stronger

discourse model. Second, the coherence relations that support the goal effect are those that

continue the narrative, or describe the outcome of the first event. Thus, the temporal succession

of utterances matches the temporal succession of events. By contrast, implicit causality effects

are supported by pre-event information about the cause, which may be harder to retrieve.

Temporal breaks in the discourse content have been associated with a drop in the use of pronouns (McCoy & Strube, 1999; Vonk, Hustinx & Simons, 1992). Third, it may be that the experiencer (i.e., the one who is not the implicit cause) is particularly salient in the discourse, because these verbs reflect the mental state of the experiencer. These observations suggest that more work is needed to understand more about how thematic role effects vary across different verbtypes.

Another possibility is that detecting thematic role effects requires specific experimental conditions. If the features of the task elicit performance that is at ceiling (~100% pronoun use) or floor (~0% pronoun use), the effect cannot be detected. Thus, in order to test any effect – including thematic roles -- it is important to find the "sweet spot" of variability in a particular task. In our study, we helped achieve the right balance in variability by adopting stricter exclusion criteria for participants than previous studies, including only participants who used some variation in their expressions. This is especially important in the increasingly popular "crowd-sourcing" approach to data collection, using Amazon Mechanical Turk. With this approach (as we used for Experiments 2 and 3), we argue that it is important to use strict inclusion criteria in order to sample participants who are trying to follow the instructions of the task. Some of these participants may be less engaged in the study than a live subject performing in front of a live experimenter, and indeed we excluded five subjects for meaningless responses. Without these stringent criteria, our sample could easily have contained low-quality data. We examined how our analyses would be affected without our selection criteria, by taking a sample of the first 20 participants in Experiment 2, regardless of whether they exhibited variation. In this sample, the thematic role effect disappeared. Nevertheless, we are confident that our

selection criteria did not create a spurious effect, in that additional analyses confirmed that the effects were present even when all participants were included.

What mechanism underlies thematic role effects?

The empirical results of this paper establish that speakers use pronouns more often for goals than sources. This question is theoretically interesting because thematic roles can influence the likelihood of the speaker mentioning the character in the subsequent utterance – that is, thematic roles affect referential predictability. The predictability of reference to goal characters is not contested. Numerous authors have claimed that goal arguments are more referentially predictable than source arguments, especially in a context where the following sentence is expected to provide information about a following event (e.g., Arnold, 2001; Fukumura & van Gompel, 2010; Kehler et al., 2008; Stevenson et al., 2004). Consistent with this, our rating study confirmed that the goal characters in our stimuli were chosen as the most likely to be mentioned. Further evidence comes from a companion story-completion study (Rosa, 2015), in which the choice of who to mention in the continuation was left up to the participant, and continuations overwhelmingly mentioned the goal character.

It is also important to point out that predictability is not the only determinant of reference form. In our data, there was also a strong tendency to use pronouns more for subjects than nonsubjects. Arnold (1998, 2001) has suggested that subjects are more predictable than nonsubjects, and Brennan (1995) has also suggested that speakers continue talking about referents in subject position more than other referents. However, our rating studies did not find that predictability or relatedness varied between subject and nonsubject continuations. We consider three possible interpretations of this finding.

One possibility is that subjecthood is unrelated to predictability. If so, subjecthood and thematic roles may influence referential form for different reasons. A second possibility is that predictability is calculated in a dynamic way, as the utterance unfolds. Early in the utterance, subjects may be perceived as highly predictable, and thus represented in a relatively accessible way. By the end of the utterance, the coherence relation may shift predictability toward the goal, but the accessible discourse representation remains. A third possibility is that subjecthood by itself is a weak indicator of predictability, but it tends to co-occur with other indicators of topicality in natural discourse, like repeated mention and pronominalization.

An open question is whether the goal bias results from the thematic roles themselves, or other correlated properties of the event structure or thematic structure. In our stimuli, the nonsubject goal and source characters had different syntactic properties: while the goal nonsubjects were arguments of the verbs, the source nonsubjects were optional adjuncts.[10] It is entirely possible that this syntactic pattern is systematically related to the predictability of goals. As Arnold (1998) argued, goal predictability reflects that fact that speakers have a tendency to choose communicative messages that involve goals. That is, given a transfer event, they are more likely to be interested in continuing to talk about the goal than the source. This pattern of human interest may also underlie the evolution of syntactic patterns.

The critical question is whether this explains the thematic-role-based predictability effect, or undermines it. We found that goals were both more predictable and more likely to be pronominalized, suggesting that predictability affects pronominalization. This stands in contrast to claims that the production of pronouns is unaffected by referential predictability, and is instead only driven by topicality-based features, like subjecthood (Kehler et al., 2008; Kehler &

---

[10] We thank an anonymous reviewer for pointing this out.

Rohde, 2013; Fukumura & van Gompel, 2010). That is, these studies support a strict separation between topicality constraints (which affect pronoun choice) and predictability constraints (which only affect likelihood of mention). If the goal effect could be instead explained by syntactic properties, could it be consistent with the topicality-based models? It's not clear that it would, because it is still the case that goals are more predictable. If it were shown that the syntactic structure of goals had topical properties too, it would provide further evidence against Kehler & Rohde's (2013) strict distinction between referential predictability and topicality, and would instead support the view that predictability and topicality are interrelated (e.g., Arnold, 1998, 2001; Givon, 1983; Prince, 1981; Tily & Piantadosi, 2009).

Another open question is whether the goal and source roles studied here represent unified categories of thematic roles. The utility of the labels "goal" and "source" is that they offer convenient categorizations that recognize one dimension of the semantic structure of transfer events, namely the starting and ending point of the transfer itself. Yet thematic roles have received numerous theoretical treatments. For example, on Jackendoff's (1987) scheme, entities have multiple thematic tiers, such that the goal/source distinction falls only on the thematic tier, while the action tier represents agent-patient relations, and the temporal tier marks the temporal framework. On this view, the predictability of goals may be related to systematicities on other thematic tiers. Levin (1993) instead groups verbs into classes that share similar syntactic and semantic properties. On this account, our goal-source verbs come from four classes (11.1 "send" verbs; 11.3 "bring" verbs; 13.1 "give" verbs, and 37.1 "quote" verbs), and our source-goal verbs come from a different four classes (10.5 "steal verbs", 13.5.1 "get" verbs, 13.5.2 "obtain" verbs, and 14 "learn verbs"). Other properties of these verb classes may influence their likelihood of being re-mentioned in discourse. Yet regardless of how we characterize the thematic nature of

our stimuli, the key finding we present is that goal and source arguments differ in both

predictability and the rate of pronominalization. Thus, the precise description of thematic roles

does not alter the important finding, which is that pronoun selection is conditioned on more than

just grammatical roles.

A remaining question is what processing mechanism underlies the effect of thematic role

predictability. For a reader or listener, predictability corresponds to the listener's ability to

anticipate the upcoming input. For a speaker, the implications of predictability are not as

straightforward. Speakers plan their utterances ahead of time, which means that they do not need

to "predict" their utterances per se.

One hypothesis we considered was that predictability in production corresponds to the

ease of planning. Information that is predictable to comprehenders is also redundant with the

context. This may support the production processes required to conceptually plan an utterance

and formulate it linguistically. If processing facilitation leads to fluency, it may support the use

of reduced expressions (Arnold, in press; Arnold & Watson, 2015). Consistent with this idea,

speakers in Experiment 1 initiated their utterances more quickly in goal continuations. However,

if utterance planning is the primary determinant of reduced reference forms, we might also

expect to see a relationship between onset latency and the production of pronouns. Yet we found

no such relationship, suggesting that planning time does not explain the thematic role effect.

Another possibility is that predictability affects reference form by directing the attention

of discourse participants toward particular referents. When something is predictable, speakers

instantiate a more-accessible representation of that referent in their mental model. This helps

speakers achieve the goal of cooperative communication, by anticipating what the addressee

considers predictable and thus accessible. Thus, it seems likely that goal continuations are

conceptually more integrated with the prior context than source continuations. This is consistent

with our finding that the goal pictures were rated as more related to the context than the source

pictures. If predictability influences referential form by affecting conceptual accessibility, such a

mechanism would be broadly consistent with claims that salience in the discourse context

increases the likelihood of using reduced expressions (Ariel, 1990; 2001; Chafe, 1976, 1994;

Gundel et al., 1993). This interpretation also predicts that thematic role effects should be

modulated by other factors that affect discourse representations, which may differ across

verbtypes.

Conclusion

In conclusion, the current study has found that thematic roles do play a role in

determining referential form, in that speakers used reduced forms for reference to goal characters

more than source characters. This empirical finding moves the field forward, by contributing to a

debate about whether thematic roles and predictability affect reference form. We established that

thematic roles do matter, at least for transfer verbs with narrative continuations. We also

confirmed that goal continuations were perceived as more predictable than sources. This

suggests that predictability can indeed affect reference form, contrary to claims that it does not.

Author Note

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the

   domain of subsequent reference. *Cognition, 73*(3), 247-264. doi:10.1016/S0010-

   0277(99)00059-1

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

Ariel, M. (2001). Accessibility theory: An overview. In T. Sanders, J. Schilperoord, & W.

   Spooren (Eds.), *Text representation: Linguistic and psycholinguistic aspects* (pp. 29– 87).

   Amsterdam: Benjamins.

Arnold, J. E. (1998). Reference form and discourse patterns (Doctoral dissertation, Stanford

   University, 1998). Dissertation Abstracts International, 59, 2950.

Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference

   continuation. *Discourse Processes, 3*, 137-162. doi:10.1207/S15326950DP3102_02

Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented

   processes. *Language and Cognitive Processes, 23*, 495-527.

   doi:10.1080/01690960801920099

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

   Compass, 4*(4), 187-203. doi:10.1111/j.1749-818X.2010.00193.x

Arnold, J. E. (2013). What should a theory of pronoun interpretation look like? Commentary of

   Kehler & Rodhe (2013): A probabilistic reconciliation of coherence-driven and

centering-driven theories of pronoun interpretation. *Theoretical Linguistics, 39*, 59-73.

doi: 10.1515/tl-2013-0003

Arnold, J.E. (2013). What should a theory of pronoun interpretation look like? Commentary on

Kehler & Rohde, "A probabilistic reconciliation of coherence-driven and centering-

driven theories of pronoun interpretation". *Theoretical Linguistics 39,* 59-73.

Arnold, J. E. (in press). Explicit and emergent mechanisms of information status. *Topics in

Cognitive Science*.

Arnold, J. E., & Griffin, Z. M. (2007). The effect of additional characters on choice of referring

expression: Everyone counts. *Journal of Memory and Language, 56*, 521-536.

doi:10.1016/j.jml.2006.09.007

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*, 31-56.

doi:10.1177/00238309040470010201

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on

durations of content and function words in conversational English. *Journal Of Memory

And Language*, *60*, 92-111. doi:10.1016/j.jml.2008.06.003

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*, 1001-1024.

doi:10.1121/1.1534836

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

137-167. doi:10.1080/01690969508407091

Brown, R., & Fish, D. (1983). The psychological causality implicit in language. *Cognition,*

    14(3), 237-273. doi:10.1016/0010-0277(83)90006-9

Buschkuehl, M., Hernandez-Garcia, L., Jaeggi, S. M., Bernard, J. A., & Jonides, J. (2014).

    Neural effects of short-term training on working memory. *Cognitive, Affective &*

    Behavioral Neuroscience, 14, 147-160. doi:10.3758/s13415-013-0244-9

Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view.

    In Charles N. Li (Ed.), *Subject and topic,* (pp. 25–56). New York: Academic Press Inc.

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

Ehrlich, K. (1980). Comprehension of pronouns. *The Quarterly Journal of Experimental*

    Psychology, *32*, 247-255. doi: 10.1080/14640748008401161

Filmer, H. L., Mattingley, J. B., & Dux, P. E. (2015). Object substitution masking for an

    attended and foveated target. *Journal of Experimental Psychology: Human Perception*

    and Performance, 41, 6-10. doi:10.1037/xhp0000024

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during

    syntactic comprehension. *PLoS ONE*, *8*, e77661. doi:10.1371/journal.pone.0077661

Frank, A. F. & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an

    optimal strategy for language production. In *Proceedings of the 30th annual conference*

    of the cognitive science society, (pp. 933–938).

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. doi:10.1037/a0031215

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.

    doi:10.1016/j.jml.2009.09.001

Gahl, S., & Garnsey, S. M. (2004). Knowledge of grammar, knowledge of usage: Syntactic

    probabilities affect pronunciation variation. *Language, 80*(4), 748-775.

    doi:10.1353/lan.2004.0185

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.

Gregory, M. L., Raymond, W. D., Bell, A., Fosler-Lussier, E., & Jurafsky, D. (1999). The effects

    of collocational strength and contextual predictability in lexical production. *Chicago*

    Linguistic Society 35, 151-166.

Gundel, J., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring

    expressions in discourse. *Language, 69*, 274-307. doi: 10.2307/416535

Hale, J. (2001, June). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings*

    of the second meeting of the North American Chapter of the Association for

    Computational Linguistics on Language technologies (pp. 1-8). Association for

    Computational Linguistics.

Hartshorne, J.K., O'Donnell, T.J., Tenenbaum, J.B. (2015). The causes and consequences

    explicit in verbs. *Language, Cognition & Neuroscience, 30*, 716-734.

    doi:10.1080/23273798.2015.1008524

Jackendoff, R. (1987). The Status of thematic relations in linguistic theory. *Linguistic Inquiry*,

    18, 369-411.

Jaeger, T. F. (2006). *Redundancy and syntactic reduction in spontaneous speech* (Doctoral

    dissertation, Stanford University). Retrieved from Proquest.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation.

    Cognitive Science, 20, 137–194. doi:10.1016/S0364-0213(99)80005-6

Jurafsky, D., Bell, A., Fosler-Lussier, E., Girand, C., & Raymond, W. D. (1998). Reduction of

    English function words in Switchboard. In *Proceedings of ICSLP-98*, (pp. 3111-3114).

    Sydney, Australia: Australian Speech Science and Technology Association, Incorporated.

Kahn, J. M., & Arnold, J. E. (2012). A processing-centered look at the contribution of givenness

    to durational reduction. *Journal of Memory and Language, 67*, 311-325.

    doi:10.1016/j.jml.2012.07.002

Kaiser, E., Li, D., Holsinger, E. (2011). Exploring the lexical and acoustic consequences of

    referential predictability. In I. Hendricks, A. Branco, S. Lalitha Devi, & R. Mitkov. (Eds.)

    Anaphora Processing and Applications, Lecture Notes in Artificial Intelligence, (Vol.

    7099, pp. 171-183). Heidelberg: Springer.

Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. CSLI Publications,

    Stanford, CA.

Kehler, A., & Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and

    centering-driven theories of pronoun interpretation. *Theoretical Linguistics, 39*, 1-37. doi:

    10.1515/tl-2013-0001.

Kehler, A., & Rohde, H. (2015). Pronominal Reference and Pragmatic Enrichment: A Bayesian

Account. In *Proceedings of the 37th Annual Conference of the Cognitive Science Society,*

(pp. 1063-1068).

Kehler, A., Kertz, L., Rohde, H., Elman, J.L. (2008). Coherence and coreference revisited.

Journal of Semantics, 25, 1-44. doi: 10.1093/jos/ffm018

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

Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic

reduction. In B. Schölkopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information

processing systems* (pp. 849–856). Cambridge, MA: MIT Press.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and

perception of speech. *Language and Speech, 6*, 172-187.

doi: 10.1177/002383096300600306

McCoy, K. F., & Strube, M. (1999). Taking time to structure discourse: Pronoun generation

beyond accessibility. In M. Hahn & S. Stoness (Eds.), *Proceedings of the twenty-first

annual conference of the Cognitive Science Society* (pp. 378–383). Mahwah, NJ:

Lawrence Erlbaum.

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*, 3526–3529. doi:10.1073/pnas.1012551108

Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.). *Radical

pragmatics*, (pp. 223–56). New York: Academic Press.

Rohde, H. (2008). *Coherence-driven effects in sentence and discourse processing*. (Doctoral

    dissertation). UCSD, San Diego, CA.

Rohde, H., & Kehler, A. (2014). Grammatical and information-structural influences on pronoun

    production. *Language, Cognition, and Neuroscience*, *29*(8), 1-16.

    doi:10.1080/01690965.2013.854918

Rohde, H., Kehler, A., & Elman, J. (2007). Pronoun interpretation as a side effect of discourse

    coherence. In *Proceedings of the 29th annual conference of the Cognitive Science Society*

    (pp. 617-622). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Rosa, E.C. (2015). *Semantic role predictability affects reference form* (Doctoral dissertation).

    University of North Carolina, Chapel Hill, NC. Retrieved from Carolina digital

    repository.

Seidenberg, M.S. & MacDonald, M.C. (1999). A probabilistic constraints approach to language

    acquisition and processing. *Cognitive Science, 23*, 569-588.

    doi:10.1207/s15516709cog2304_8

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.

    doi:10.1016/j.specom.2005.06.005

Staub, A. & Clifton Jr., C. (2006). Syntactic prediction in language comprehension: Evidence

    from either…or. *Journal of Experimental Psychology: Learning, Memory, and*

    Cognition, 32, 425-436. doi:10.1037/0278-7393.32.2.425

Stevenson, R.J., Crawley, R.A., Kleinman, D. (1994). Thematic roles, focus, and the

representation of events. *Language and Cognitive Processes, 9*, 519-548.

doi:10.1080/01690969408402130

Tily, H., & Piantadosi, S. (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. Amsterdam, The Netherlands.

van Berkum, J.J., Brown, C.M., Zwitserlood, P., Kooijman, V., Hagoort, P. (2005). Anticipating

upcoming words in discourse: evidence from ERPs and reading times. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 31, 443-467.

doi:10.1037/0278-7393.31.3.443

van Rij. J., van Rijn, H., & Hendriks, P. (2012). How WM load influences pronoun

interpretation. In N. Rußwinkel, U. Drewitz & H. van Rijn (Eds.), *Proceedings of the*

11th International Conference on Cognitive Modeling, Berlin: Universitaetsverlag der

TU Berlin.

Vonk, W., Hustinx, L. G., & Simons, W. H. (1992). The use of referential expressions in

structuring discourse. Special Issue: Discourse representation and text processing.

Language and Cognitive Processes, 7, 301–333. doi:10.1080/01690969208409389

Zipf, G. (1936). *The Psychobiology of Languag*e. London: Routledge.

Zerkle, S., Rosa, E. C., & Arnold, J. E. (under review). *Thematic role predictability and planning*

affect word duration. Manuscript submitted for publication.

Zerkle, S., Rosa, E. C., & Arnold, J. E. (2015, March). Do addressee gestures influence

    predictability of spoken reference form? Poster session presented at the CUNY

    Conference on Human Sentence Processing, Los Angeles, CA.